

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

March 19, 2002, 09:34:56 ; Search time 1179.99 Seconds
(without alignments)
1120.120 Million cell updates/sec

Run on:

Title: US-09-684-016-48411

Perfect score: 123
Sequence: 1 caggtatgttaatgtgaga.....tcaattatccacacacaa 123

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size: 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

EST:
1: em_estfun:
2: em_esthum:
3: em_estcom:
4: em_estcom:
5: em_estpl:
6: em_estlba:
7: em_estro:
8: em_estro:
9: em_hic:
10: gb_est1:
11: gb_est2:
12: gb_hic:
13: gb_gss:
14: em_gss_fun:
15: em_gss_hum:
16: em_gss_inv:
17: em_gss_pln:
18: em_gss_pro:
19: em_gss_rtd:
20: em_gss_vrl:
21: em_gss_other:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	36.6	491	11	BG551127 sad33f01.
2	44	35.8	497	11	BG5511208 sad34e05.
3	44	35.8	530	10	AM761136 s163f12.Y
4	44	35.8	764	10	AM349333 GM210007A
5	29	23.6	540	10	AM432561 sh75h06.Y
6	26	21.1	765	10	BE659470 CM700009B
7	21	17.1	349	13	BE659470 HS_5183.A
8	21	17.1	457	11	BE999004 EST430727
9	21	17.1	514	10	AM586761 EST318384
10	21	17.1	539	11	BE650924 NF098E01E
11	21	17.1	574	10	BE124753 EST333788
12	21	17.1	586	11	BE997991 EST429714

C 13	21	17.1	635	10	AM584723	AM584723 N210838e
C 14	21	17.1	655	10	AM584724	AM584724 N210839e
C 15	21	17.1	682	10	AM584948	AM584948 N211110e
C 16	18	14.6	444	13	AO597145	AO597145 HS_5240.A
C 17	18	14.6	468	10	BE653345	BE653345 UI-M-AL1-
C 18	18	14.6	487	13	AO597171	AO597171 HS_5240.A
C 19	18	14.6	533	13	BH054207	BH054207 RPT1-24-3
C 20	18	14.6	574	13	A2290583	A2290583 RPT1-23-1
C 21	18	14.6	595	13	A2243191	A2243191 RPT1-23-3
C 22	18	14.6	599	11	BR815581	BR815581 MR2-C1012
C 23	18	14.6	606	13	A2396970	A2396970 IM0161D13
C 24	18	14.6	634	11	BE208765	BE208765 601872235
C 25	17	13.8	167	10	AV326766	AV326766 AV326766
C 26	17	13.8	211	10	AA874072	AA874072 vW88b10.X
C 27	17	13.8	249	10	AV337137	AV337137 AV337137
C 28	17	13.8	277	10	BB82343	BB82343 BB82343
C 29	17	13.8	286	10	AV277980	AV277980 AV277980
C 30	17	13.8	286	10	BB248570	BB248570 BB248570
C 31	17	13.8	301	10	A1466360	A1466360 vW88b10.Y
C 32	17	13.8	316	11	B118737	B118737 d2904fs.X
C 33	17	13.8	329	11	B1266425	B1266425 NF097C04I
C 34	17	13.8	334	10	AA244200	AA244200 nc06c06.s
C 35	17	13.8	360	11	C70635	C70635 C70635 Yui1
C 36	17	13.8	366	11	B1266362	B1266362 NF092D09I
C 37	17	13.8	450	10	AM457639	AM457639 UI-M-BH3-
C 38	17	13.8	456	11	BE649642	BE649642 NF082D01E
C 39	17	13.8	489	11	BE989827	BE989827 UI-M-B21-
C 40	17	13.8	519	10	AM394535	AM394535 sb32c05.Y
C 41	17	13.8	526	13	A2626769	A2626769 OM14.H0
C 42	17	13.8	544	11	BG048538	BG048538 CMT14.H0
C 43	17	13.8	541	13	A2961481	A2961481 RPT1-23-2
C 44	17	13.8	555	13	A2561989	A2561989 RPT1-23-2
C 45	17	13.8	557	13	A2149056	A2149056 SP_0046.A

ALIGNMENTS

RESULT 1
LOCUS BG551127
DEFINITION sad33f01.y1 Gm-cl074 Glycine max cDNA clone GENOME SYSTEMS CLONE
ID: Gm-cl074-2162 5' similar to TR:Q9SV14 Q9SV14 ES43 LIKE PROTEIN
'' mRNA sequence.
ACCESSION BG551127
VERSION BG551127.1
KEYWORDS GI:13562907
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.

REFERENCE
AUTHORS 1 (bases 1 to 491)
Shoemaker, R., Keim, P., Vodka, L., Erpelidg, J., Coryell, V., Rhanna
, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
, T., Person, B., Swaller, T., Gibbons, M., Page, D., Harvey, N., Schurk
, R., Rutter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
, R., Waterston, R. and Wilson, R.

TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Public Soybean EST Project
Contact: Shoemaker R/Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1810
Fax: 314 286 1810
Email: est@wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or

FEATURES

info@genomesystems.com web site: www.genomesystems.com
High quality sequence stop: 421.
Location/Qualifiers

1..491

/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1074-2162"
/clone_1lb="Gm-c1074"
/tissue_type="seedlings induced for HR (hypersensitive response)"
/dev_stage="9-11 day old"
/lab_host="DH10B"

/note="Vector: Bluescript II SK(+)

Site:1: EcoRI; Site:2: XhoI. The cDNA library was constructed from mRNA isolated from 9-11 day old seedlings that were induced for HR (hypersensitive response) by vacuum infiltrating plant tissue with *Pseudomonas syringae* pv. *glycinea* carrying the *avrB* gene (Genetics 141:1597-1604). Plant tissue (expanded unifoliate leaves) was collected at 2, 4, 8, 12, 24, 36, and 53 hrs after inoculation and their mRNA pooled equally for cDNA construction. The library was prepared using the Stratagene Bluescript II SK(+) library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with an XhoI restriction site. EcoRI adaptors were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA insert is protected from XhoI digestion via methylation during first strand synthesis. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the Bluescript vector. The ligated cDNA fragments were transformed into *E. coli* Electromax DH10B host cells. Plant care, inoculations, and library construction were performed by Steve Clough (Lila Vodkin lab, University of Illinois).

BASE COUNT 117 a 138 c 129 g 107 t
ORIGIN

Query Match 36.6%; Score 45; DB 11; Length 491;
Best local Similarity 100.0%; Pred. No. 4.9e-14;

Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 ggtattgtaaatgtgagatgcttaacaacctgagcctaattgg 47
|||||
Db 430 GGTATTGTAATGTGAGATGCTTACCAACCTGATGACCTAATGG 474

RESULT 2

BG551208

LOCUS

DEFINITION

bg551208 497 bp mRNA EST 09-APR-2001
sad34e05.y1 Gm-c1074 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1074-2145 5' similar to TR:O5SV14 O9SV14 ES43 LIKE PROTEIN ; mRNA sequence.

ACCESSION BG551208 GI:13562988
VERSION BG551208.1 GI:13562988
KEYWORDS EST.

SOURCE

ORGANISM

soybean.
Glycine max
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE 1 (bases 1 to 497)
AUTHORS

Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Stepien, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Rutter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., Mccann, R., Waterston, R. and Wilson, R.

TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project

FEATURES

source

Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
High quality sequence stop: 380.
Location/Qualifiers

1..497

/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1074-2145"
/clone_1lb="Gm-c1074"
/tissue_type="seedlings induced for HR (hypersensitive response)"
/dev_stage="9-11 day old"
/lab_host="DH10B"

/note="Vector: Bluescript II SK(+); Site:1: EcoRI; Site:2: XhoI. The cDNA library was constructed from mRNA isolated from 9-11 day old seedlings that were induced for HR (hypersensitive response) by vacuum infiltrating plant tissue with *Pseudomonas syringae* pv. *glycinea* carrying the *avrB* gene (Genetics 141:1597-1604). Plant tissue (expanded unifoliate leaves) was collected at 2, 4, 8, 12, 24, 36, and 53 hrs after inoculation and their mRNA pooled equally for cDNA construction. The library was prepared using the Stratagene Bluescript II SK(+) library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with an XhoI restriction site. EcoRI adaptors were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA insert is protected from XhoI digestion via methylation during first strand synthesis. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the Bluescript vector. The ligated cDNA fragments were transformed into *E. coli* Electromax DH10B host cells. Plant care, inoculations, and library construction were performed by Steve Clough (Lila Vodkin lab, University of Illinois).

BASE COUNT 119 a 140 c 129 g 109 t
ORIGIN

Query Match 35.8%; Score 44; DB 11; Length 497;
Best local Similarity 100.0%; Pred. No. 1.7e-13;

Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gttattgtaaatgtgagatgcttaacaacctgagcctaattgg 47
|||||
Db 431 GTATTGTAATGTGAGATGCTTACCAACCTGATGACCTAATGG 474

RESULT 3

AW761136

LOCUS

DEFINITION

aw761136 530 bp mRNA EST 21-NOV-2000
sl63fl2.y1 Gm-c1077 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-c1027-6024 5' similar to TR:O65462 O65462 RECEPTOR LIKE PROTEIN ; mRNA sequence.

ACCESSION AW761136 GI:7693038
VERSION AW761136.1 GI:7693038
KEYWORDS EST.

SOURCE

ORGANISM

soybean.
Glycine max
Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae: eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE 1 (bases 1 to 530)
AUTHORS Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna

TITLE
JOURNAL
COMMENT

A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, R., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schuck, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R./Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Insert Length: 1680 Std Error: 0.00
High quality sequence stop: 420.

FEATURES
Source

1. 530
Location/Qualifiers
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl027-6024"
/clone.lib="Gm-cl027"
/tissue_type="cotyledons of 3- and 7-day-old Williams seedlings"
/lab_host="PH10B"
/note="Vector: pBluescript II SK+; Site-1: EcoRI; Site-2: XhoI; This cDNA library was constructed from mRNA isolated from cotyledons of 3- and 7-day-old Williams seedlings which were propagated on paper towels with distilled water. The cotyledons were flash-frozen in liquid nitrogen, then lyophilized for 72 hours. Unequal amounts of mRNA was used for cDNA synthesis. Stratagene's cDNA Synthesis Kit (catalog number 200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA was hemimethylated. A modification of Stratagene's first-strand synthesis primer was used. An anchor nucleotide (V-A, C, or G) was added to the 3' end of the primer (GAGAGAGAGAGAGAGACTGCTCGAG(T)18) to anchor the primer at the 5' end of the poly(A) tract. After second-strand synthesis, the cDNA ends were filled in with cloned Pfu DNA, ligated to EcoRI adapters and subsequently phosphorylated. The XhoI site within the first-strand synthesis primer was then restricted by digestion with XhoI; all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500 bp cutoff, using GibcoBRL Life Technologies' cDNA Size Fractionation column. The column eluent was then ligated into Stratagene's pBluescript(tm) II XR predigested vector (pBluescript II SK(+)) that has been digested with EcoRI and XhoI, and phosphorylated by Stratagene). 97% of the white and blue colonies appear to contain recombinant plasmids with cDNA inserts, based on size (n=30). This library was constructed by Dr. Paul Keim and Dr. Virginia Coryell."

BASE COUNT
ORIGIN

136 a 145 c 136 g 113 t
Query Match 35.8%; Score 44; DB 10; Length 530;
Best Local Similarity 100.0%; Pred. No. 1.7e-13;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 gttattgaatgtgagatgccttaacaacctgattgacctaatgg 47
|||||
DB 407 GTATTGTAATGTGAGATGCTTACCAACCTGATGACCTAATGG 450

RESULT 4
AM349333/C
LOCUS
DEFINITION
AM349333 764 bp mRNA EST 04-OCT-2000
GM210007/A20A6 Gm-r1021 Glycine max cDNA clone Gm-r1021-2316 3',
rRNA sequence.
ACCESSION
AM349333
VERSION
AM349333.1 GI:6847043
KEYWORDS
EST.
SOURCE
soybean.
ORGANISM
Glycine max

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 764)
Vodkin, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Coryell, V., Erpelting, J., Rapp, C., Shoop, E., Pardinas, J., Liu, L. and Lewin, H. A Functional Genomics Program for Soybean (NSF 9872565)
Unpublished (1999)
Other ESTs: AF461129
Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu

FEATURES
source

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.
Location/Qualifiers
1. 764
/organism="Glycine max"
/cultivar="Williams"
/db_xref="taxon:3847"
/clone="Gm-r1021-2316"
/clone.lib="Gm-r1021"
/tissue_type="root"
/lab_host="XL10-GOLD"
/note="Vector: pBluescript II XR; Site-1: EcoRI; Site-2: XhoI; Library Gm-r1021 is a sequence-driven, rerecked set of the original library Gm-cl004 which was prepared from root cDNA. The mRNA was isolated from entire roots of 8 day old 'Williams' seedlings which were propagated on paper towels with distilled water. Stratagene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. The Gm-cl004 library was constructed by Dr. Paul Keim & Virginia H. Coryell, Department of Biology, Box5640, Northern Arizona University, Flagstaff, AZ 86011, email: paul.keim@nau.edu, virginia.coryell@nau.edu. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Computational Biology Centers, University of Minnesota,
http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
Rerecking was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and sequencing by the Keck Center for Comparative and Functional Genomics,
http://www.life.uiuc.edu/biotech/keck.html."

BASE COUNT
ORIGIN

239 a 161 c 151 g 193 t 20 others

Query Match 35.8%; Score 44; DB 10; Length 764;
Best Local Similarity 100.0%; Pred. No. 1.8e-13;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 gtaattgtaaatgtgagatgccttacacccctgatacctaagg 47
 ||||||||||||||||||||||||||||||||||||||||||||
 Db 566 GTATTGTAATGTGAGATGCTTACCAACCTGACTGACTAATGG 523

RESULT 5
 AM432561 540 bp mRNA EST 18-JUL-2000
 LOCUS AM432561
 DEFINITION sh75h06.y1 Gm-cl015 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl015-5652 5' similar to TR:0654462 0654462 RECEPTOR LIKE PROTEIN ; mRNA sequence.
 AM432561
 ACCESSION AM432561.1 GI:6963868
 VERSION EST.
 KEYWORDS soybean.
 SOURCE Glycine max
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE 1 (bases 1 to 540)
 AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpellding,J., Coryell,V., Khanna,A., Bolla,B., Merrig,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Thelming,B., Allen,M., Bowers,I., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shih,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
 TITLE Public Soybean EST Project
 JOURNAL Unpublished (1999)
 COMMENT Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@wustl.edu
 This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
 Insert length: 1008 Std Error: 0.00
 Seq primer: -40RP from Gibco
 High quality sequence stop: 380.
 Location/Qualifiers
 1..540
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl015-5652"
 /clone_lib="Gm-cl015"
 /tissue_type="Mature flowers, field grown plants"
 /lab_host="XL10-Gold"
 /note="Vector: pBluescript II XR; Site1: EcoRI; Site2: XhoI; This cDNA library was constructed from mRNA isolated from mature flowers of field grown plants. The cDNA library was prepared using the Stratagene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into XL10-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpellding."

BASE COUNT 123 a 142 c 138 g 137 t

Query Match 23.6%; Score 29; DB 10; Length 540;
 Best Local Similarity 100.0%; Pred. No. 2.5e-05;
 Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 4 gtaattgtaaatgtgagatgccttacacac 32
 ||||||||||||||||||||||||||||||||||||||||||||
 Db 468 GTATTGTAATGTGAGATGCTTACCAACC 496

RESULT 6
 BE659470 765 bp mRNA EST 24-MAY-2001
 LOCUS BE659470/c
 DEFINITION Gm700009B10H8 Gm-r1070 glycine max cDNA clone Gm-r1070-3447 3', mRNA sequence.
 BE659470
 ACCESSION BE659470.1 GI:9985362
 VERSION EST.
 KEYWORDS soybean.
 SOURCE Glycine max
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE 1 (bases 1 to 765)
 AUTHORS Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V., Erpellding,J., Rapp,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
 TITLE A Functional Genomics Program for Soybean (NSF 9872565)
 JOURNAL Unpublished (1999)
 COMMENT Other ESTs: AM432561 corresponding to Gm-cl015-5652 (5')
 Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)
 Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics
 University of Illinois
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
 Tel: (217) 244-6147
 Fax: (217) 333-4582
 Email: l-vodkin@uiuc.edu
 This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
 Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.
 Location/Qualifiers
 1..765
 /organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="Gm-r1070-3447"
 /clone_lib="Gm-r1070"
 /note="The library Gm-r1070 is a sequence-driven, rerecked set of 9,216 clones selected from cDNA libraries from various tissues and stages of development of soybean that represent 2,639 sequences from immature cotyledons, 1,770 from immature seed coats, 3,938 from flowers, and 869 from young pods. The 5' ESTs of the source clones from the different libraries was used to select singletons, or a representative of each contig, which were rerecked to form library Gm-r1070. The cDNA clones of the rerecked Gm-r1070 library were then sequenced at the 3' end. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota, http://www.cbc.umn.edu/research/Projects/Soybean/Index.html. Rerecking was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and 3' sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, http://www.life.uiuc.edu/biotech/keck.html. Note: The corresponding 5' EST from each clone in the Gm-r1070 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST.'"

BASE COUNT 227 a 167 c 150 g 197 t 24 others

ORIGIN

Query Match 21.1%: Score 26; DB 10; Length 765;
 Best Local Similarity 100.0%; Pred. No. 0.0011;
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 49 gcaatgcagagcgcactgactgct 74
 ||||||||||||||||||
 Db 497 GCAATGCAGAGCGCTGCACTGACTGCT 472

RESULT 7
 A0451805/c 349 bp DNA GSS 21-APR-1999
 LOCUS HS.5183.A1.B08.T7A.RPCR-11 Human Male BAC Library Homo sapiens
 DEFINITION genomic clone Plate=759 Col-15 Row-C, DNA sequence.
 ACCESSION A0451805
 VERSION A0451805.1 GI:4592975
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 349)
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
 Mahairas,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
 Hood,L.
 Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 CONTACT: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC
 library availability, please contact Plier de Jong
 (pielerdejong.med.bufileo.edu). Clones may be purchased from
 BACPAC Resources (http://bacpac.med.bufileo.edu/ordering_bac.htm)
 or from Resear h Genetics (info@resgen.com). BAC end Web Server:
 http://www.htsc.washington.edu
 Plate: 759 row: C column: 15
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 349.

FEATURES

source
 1. 349
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=759 Col=15 Row=C"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="Male"
 /note="Vector: pBACe3.6; Site.1: EcoRI; Site.2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor
 and partially digested with a combination of EcoRI and
 EcoRI Methylase. Size selected DNA was cloned into the
 pBACe3.6 vector at EcoRI sites"
 BASE COUNT 83 a 79 c 86 g 101 t
 ORIGIN

Query Match 17.1%: Score 21; DB 13; Length 349;
 Best Local Similarity 100.0%; Pred. No. 0.53;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 ttatctcaattatccaca 118
 ||||||||||||||||||
 Db 69 TTATCTCAATTATTCACA 49

RESULT 8

BE999004 457 bp mRNA EST 06-OCT-2000
 LOCUS EST430727 GVSN Medicago truncatula cDNA clone pgvsn-13016, mRNA
 DEFINITION sequence.
 ACCESSION BE999004
 VERSION BE999004.1 GI:10699280
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
 Medicago.

REFERENCE 1 (bases 1 to 457)
 Fedorova,M., Pierson,B.L., Samac,D.A., Gantt,J.S., Vance,C.P., Town
 ,C.D., Bowman,C.L., Craven,M.B., Cho,J., and Fraser,C.M.
 ESTs from senescent nodules of Medicago truncatula
 Unpublished (2000)
 CONTACT: Carol P. Vance
 Department of Agronomy and Plant Genetics
 University of Minnesota
 411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
 Tel: 612 625 5715
 Fax: 651-649-5058
 Email: Vance004@maroon.tc.umn.edu
 University of Minnesota name: M272982e TIGR sequence name:
 MTKMX20TK More information is available at:
 http://chryslie.tamu.edu/medicago
 Seq primer: SKmod (CTA GAA CTA gtc gat CC).
 FEATURES
 source
 1. 457
 Location/Qualifiers
 /organism="Medicago truncatula"
 /cultivar="genotype A17"
 /db_xref="taxon:3880"
 /clone="pgvsn-13016"
 /clone_lib="GVSN"
 /rissue_type="senescent root nodules"
 /dev_stage="mixture of effective nodules from 40 day old
 plants harvested 36 hours post shoot removal and nodules
 collected from 2 month old plants at mid-pod stage"
 /lab_host="E. coli strain SOLR"
 /note="Vector: pBluescript SK +/-; Site.1: EcoRI; Site.2:
 XhoI; cDNA was prepared from polyA+ enriched RNA from the
 mixture of effective nodules of 40 day old plants
 harvested 36 hours post shoot removal and nodules
 collected from 2 month old plants at mid-pod stage. The
 cDNA was directionally ligated into the Uni-ZAP XR vector
 from stratagene and packaged using Gigapack III Gold
 packaging extracts. Plasmids containing cDNA inserts were
 excised from the recombinant lambda-ZAP phage using
 Ex-assist helper phage and propagated in SOLR cells."

BASE COUNT 126 a 90 c 109 g 132 t
 ORIGIN

Query Match 17.1%: Score 21; DB 11; Length 457;
 Best Local Similarity 100.0%; Pred. No. 0.54;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 gtattgtaattgagatgcc 24
 ||||||||||||||||||
 Db 416 GTATTGTAATGTGAGATGCC 436

RESULT 9
 AW586761 514 bp mRNA EST 07-SEP-2000
 LOCUS EST318384 MRAM Medicago truncatula/c10mus versiforme mixed EST
 DEFINITION library cDNA clone pMAM-55G12, mRNA sequence.
 ACCESSION AW586761
 VERSION AW586761.1 GI:7266275

KEYWORDS EST.
SOURCE Medicago truncatula/Glomus versiforme mixed EST library.
ORGANISM Medicago truncatula/Glomus versiforme mixed EST library.
REFERENCE Eukaryota, mixed EST libraries.
AUTHORS 1 (bases 1 to 514)
Harrison, M.J., Liu, J., Peng, H., Gonzales, M., Ellis, L., Town, C.D.,
Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E., and Fraser, C.M.
ESTs from roots of Medicago truncatula after colonization with
Glomus versiforme
JOURNAL Unpublished (2000)
COMMENT Contact: Maria J. Harrison
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73401, USA
Tel: 580-223-5810
Fax: 580-221-7380
Email: mjharrison@noble.org
Noble EST name: N255126
TIGR sequence name: WTD4V42PK
More information is available at:
'http://chryslie.tamu.edu/medicago/'
Seq primer: SKmod (CTA gAA CTA gAT CC).
Location/Qualifiers
1. .514
/organism="Medicago truncatula/Glomus versiforme mixed EST
library"
/cultivar="Medicago truncatula genotype A17"
/db_xref="taxon:119092"
/clone="pMHAM-55G12"
/clone_lib="MHAM"
/tissue_type="roots colonized with Glomus versiforme"
/dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days
post-inoculation with Glomus versiforme. The library was
made from a mixture of RNA from each of these stages."
/lab_host="E. coli strain XLOLR"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA from
roots harvested at 10, 17, 22, 31 and 38 days
post-inoculation with Glomus versiforme. The cDNA was
directionally ligated into the Uniap XR vector from
stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in XLOLR cells."

BASE COUNT 135 a 101 c 127 g 151 t
ORIGIN

Query Match 17.1%; Score 21; DB 10; Length 514;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gtattgtaaatgtgagatgcc 24
|||||
Db 415 GATTGTAATGTGAGATGCC 435

RESULT 10
BF650924 539 bp mRNA EST 20-DEC-2000
LOCUS NF098E01EC1F1005 Elicited cell culture Medicago truncatula cDNA
DEFINITION clone NF098E01EC 5', mRNA sequence.
ACCESSION BF650924
VERSION BF650924.1 GI:11916054
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
1 (bases 1 to 539)
Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J.,

TITLE Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation -
Center for Medicago Genomics Research
JOURNAL Unpublished (2000)
COMMENT Contact: Dixon RA
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7302
Fax: 580 221 7380
Email: radixon@noble.org
Insert Length: 539 Std Error: 0.00
Insert row: E column: 01
Seq primer: TCACACAGGAACAGCTATGAC.
Location/Qualifiers
1. .539
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF098E01EC"
/clone_lib="Elicited cell culture"
/tissue_type="Cell cultures derived from root tissues"
/dev_stage="Cell suspensions were subcultured every 14
days. Cells were induced six days after subculture"
/note="Vector: Lambda Zap; Cells were induced with yeast
cell wall extracts equivalent to 50ug/ml glucose in the
final concentration. Samples were taken at 0.5, 1, 12 and
24 hours after induction. Equal amounts of RNA from each
time point were pooled and used for mRNA isolation."

BASE COUNT 146 a 108 c 125 g 157 t 3 others
ORIGIN

Query Match 17.1%; Score 21; DB 11; Length 539;
Best Local Similarity 100.0%; Pred. No. 0.55;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gtattgtaaatgtgagatgcc 24
|||||
Db 443 GATTGTAATGTGAGATGCC 463

RESULT 11
BE124753 574 bp mRNA EST 07-SEP-2000
LOCUS EST393788 GVN Medicago truncatula cDNA clone pGVN-67F7, mRNA
DEFINITION sequence.
ACCESSION BE124753
VERSION BE124753.1 GI:8529310
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
1 (bases 1 to 574)
Fedorova, M., Pierson, B.L., Samac, D.A., Vance, C.P., Gantt, J.S., Peng
H., Ellis, L., Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S.,
Holt, I.E. and Fraser, C.M.
ESTs from one month old nitrogen-fixing root nodules of Medicago
truncatula
JOURNAL Unpublished (2000)
COMMENT Contact: Carroll P. Vance
Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
Tel: 612 625 5715
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
University of Minnesota name: N261622e
TIGR sequence name: MTCC287K
More information is available at:
http://chryslie.tamu.edu/medicago

FEATURES
source
Seq primer: SKmod (CTA gAA CTA gTg gAT CC).
Location/Qualifiers
1..574

/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pGVN-67F7"
/clone.lib="GVN"
/tissue.type="N2-fixing root nodules"
/dev_stage="effective root nodules harvested one month post inoculation with Sinorhizobium meliloti"
/lab_host="E. coli strain XLOLR"
/note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2: XhoI; cDNA was prepared from polyA+ enriched RNA from effective root nodules harvested one month post inoculation with Sinorhizobium meliloti. The cDNA was directionally ligated into the Uni-ZAP XR vector from Strategene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in XLOLR cells."
BASE COUNT
155 a 114 c 137 g 167 t 1 others
ORIGIN

Query Match 17.1%; Score 21; DB 10; Length 574;
Best Local Similarity 100.0%; Pred. No. 0.55;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 gttatgtaaatgtgagatgcc 24
|||||
Db 439 GTATTGTAATGTGAGATGCC 459

RESULT 12
BE997991 586 bp mRNA EST 06-OCT-2000
LOCUS EST429714 GVSN Medicago truncatula cDNA clone pGVSN-8F9, mRNA
DEFINITION sequence.
ACCESSION BE997991
VERSION BE997991.1 GI:10698267
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.
1 (bases 1 to 586)
REFERENCE
1 Fedorova, M., Pierson, B.L., Samac, D.A., Gantt, J.S., Vance, C.P., Town, C.D., Bowman, C.L., Craven, M.B., Cho, J., and Fraser, C.M.
ESTs from senescent nodules of Medicago truncatula
Unpublished (2000)
CONTACT: Carroll P. Vance
Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
Tel: 612 625 5715
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
University of Minnesota name: M271969e TIGR sequence name: MTKRK297K More information is available at: http://chryslie.tamu.edu/medicago
Seq primer: SKmod (CTA gAA CTA gTg gAT CC).
FEATURES
source
Location/Qualifiers
1..586
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="pGVSN-8F9"
/clone.lib="GVSN"
/tissue.type="senescent root nodules"
/dev_stage="mixture of effective nodules from 40 day old

plants harvested 36 hours post shoot removal and nodules collected from 2 month old plants at mid-pod stage"
/lab_host="E. coli strain SOLR"
/note="Vector: pBluescript SK +/-; Site.1: EcoRI; Site.2: XhoI; cDNA was prepared from polyA+ enriched RNA from the mixture of effective nodules of 40 day old plants harvested 36 hours post shoot removal and nodules collected from 2 month old plants at mid-pod stage. The cDNA was directionally ligated into the Uni-ZAP XR vector from Strategene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in SOLR cells."
BASE COUNT
158 a 114 c 143 g 171 t
ORIGIN

Query Match 17.1%; Score 21; DB 11; Length 586;
Best Local Similarity 100.0%; Pred. No. 0.55;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 gttatgtaaatgtgagatgcc 24
|||||
Db 432 GTATTGTAATGTGAGATGCC 452

RESULT 13
AM584723 635 bp mRNA EST 07-SEP-2000
LOCUS N210838e MHAM Medicago truncatula/Glomus versiforme mixed EST library
DEFINITION library cDNA clone MHAM-7K19, mRNA sequence.
ACCESSION AM584723
VERSION AM584723.1 GI:7261777
KEYWORDS EST.
SOURCE Medicago truncatula/Glomus versiforme mixed EST library.
ORGANISM Medicago truncatula/Glomus versiforme mixed EST library
Eukaryota; mixed EST libraries.
REFERENCE
1 (bases 1 to 635)
AUTHORS Harrison, M.J., Liu, J., Peng, H., Gonzales, M., Ellis, L., Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E., and Fraser, C.M.
ESTs from roots of Medicago truncatula after colonization with Glomus versiforme
Unpublished (2000)
CONTACT: Harrison M.J.
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73401
Tel: 580-223-5810
Fax: 580-221-7380
Email: mjharrison@noble.org
Other name: MHAM-7a-F10; Date: 3/14/00; Updated to the Database of Expressed Sequence Tags (dbEST) on 04/27/00; More information is available at: http://chryslie.tamu.edu/medicago.
Seq primer: T3.
FEATURES
source
Location/Qualifiers
1..635
/organism="Medicago truncatula/Glomus versiforme mixed EST library"
/cultivar="Medicago truncatula genotype A17"
/db_xref="taxon:119092"
/clone="MHAM-7K19"
/clone.lib="MHAM"
/tissue.type="roots colonized with Glomus versiforme"
/dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The library was made from a mixture of RNA from each of these stages."
/lab_host="E. coli strain XLOLR"
/note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The cDNA was directionally ligated into the UniZAP XR vector from Strategene and packaged using Gigapack III Gold packaging

extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XL0LR cells."

BASE COUNT 202 a 159 c 103 g 171 t

ORIGIN

Query Match 17.1%; Score 21; DB 10; Length 635;
Best Local Similarity 100.0%; Pred. No. 0.55;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gttatgtaaatgtgagatgcc 24
|||||

Db 468 GTATTGTAATGTGAGATGCC 448

RESULT 14
AM584724/c 655 bp mRNA EST 07-SEP-2000
LOCUS
DEFINITION N210839e MHAM Medicago truncatula/Glomus versiforme mixed EST
library cDNA clone MHAM-7K21, mRNA sequence.

ACCESSION AM584724
VERSION AM584724.1 GI:7261778
KEYWORDS EST.
SOURCE Medicago truncatula/Glomus versiforme mixed EST library.
ORGANISM Medicago truncatula/Glomus versiforme mixed EST library
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 655)
Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
ESTs from roots of Medicago truncatula after colonization with Glomus versiforme
Unpublished (2000)
JOURNAL Contact: Harrison M.J.
COMMENT Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73401
Tel: 580-223-5810
Fax: 580-223-7380
Email: mharrison@noble.org
Other name: MHAM-7a-F11; Date: 3/14/00; updated to the Database of Expressed Sequence Tags (dbEST) on 04/27/00; More information is available at 'http://chrystle.tamu.edu/medicago'.
Seq primer: T3.

FEATURES

source

1.655
Location/Qualifiers
/organism="Medicago truncatula/Glomus versiforme mixed EST library"
/cultivar="Medicago truncatula genotype A17"
/db_xref="taxon:119092"
/clone="MHAM-7K21"
/clone_lib="MHAM"
/tissue_type="roots colonized with Glomus versiforme"
/dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The library was made from a mixture of RNA from each of these stages."
/lab_host="E. coli strain XL0LR"
/note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The cDNA was directionally ligated into the UniZap XR vector from Stragene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XL0LR cells."

BASE COUNT 207 a 163 c 108 g 177 t

ORIGIN

Query Match 17.1%; Score 21; DB 10; Length 655;
Best Local Similarity 100.0%; Pred. No. 0.55;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gttatgtaaatgtgagatgcc 24
|||||

Db 468 GTATTGTAATGTGAGATGCC 448

RESULT 15
AM584948 682 bp mRNA EST 07-SEP-2000
LOCUS
DEFINITION N211110e MHAM Medicago truncatula/Glomus versiforme mixed EST
library cDNA clone MHAM-7J12, mRNA sequence.

ACCESSION AM584948
VERSION AM584948.1 GI:7262002
KEYWORDS EST.
SOURCE Medicago truncatula/Glomus versiforme mixed EST library.
ORGANISM Medicago truncatula/Glomus versiforme mixed EST library
Eukaryota; mixed EST libraries.
REFERENCE 1 (bases 1 to 682)
Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
ESTs from roots of Medicago truncatula after colonization with Glomus versiforme
Unpublished (2000)
JOURNAL Contact: Harrison M.J.
COMMENT Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73401
Tel: 580-223-5810
Fax: 580-223-7380
Email: mharrison@noble.org
Other name: MHAM-7d-E06; Date: 3/14/00; updated to the Database of Expressed Sequence Tags (dbEST) on 04/27/00; More information is available at 'http://chrystle.tamu.edu/medicago'.
Seq primer: T3.

FEATURES

source

1.682
Location/Qualifiers
/organism="Medicago truncatula/Glomus versiforme mixed EST library"
/cultivar="Medicago truncatula genotype A17"
/db_xref="taxon:119092"
/clone="MHAM-7J12"
/clone_lib="MHAM"
/tissue_type="roots colonized with Glomus versiforme"
/dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The library was made from a mixture of RNA from each of these stages."
/lab_host="E. coli strain XL0LR"
/note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The cDNA was directionally ligated into the UniZap XR vector from Stragene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XL0LR cells."

BASE COUNT 212 a 172 c 115 g 183 t

ORIGIN

Query Match 17.1%; Score 21; DB 10; Length 682;
Best Local Similarity 100.0%; Pred. No. 0.56;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 gttatgtaaatgtgagatgcc 24
|||||

Db 468 GTATTGTAATGTGAGATGCC 448

RESULT 16
AO597145/c 444 bp DNA GSS 08-JUN-1999
LOCUS
DEFINITION HS-5240_A1_B11-SP6E RPCT-11 Human Male BAC Library Homo sapiens

genomic clone Plate-816 Col-21 Row-C, DNA sequence.

ACCESSION A0597145
 VERSION A0597145.1 GI:5028357
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 444)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,V., Young,V., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
 Plate: 816 row: C column: 21
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 444.

FEATURES
 source Location/Qualifiers
 1..444
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate-816 Col-21 Row-C"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"

BASE COUNT 111 a 96 c 103 g 131 t 3 others

ORIGIN

Query Match 14.6%; Score 18; DB 13; Length 444;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 ttatctcaattatctcc 115
 |||
 Db 47 TTATCTCATATTATTC 30

RESULT 17
 BE653345 468 bp mRNA EST 06-SEP-2000
 LOCUS UI-M-AL1-ah1-c-07-0-UI.r1 NIH BMAP_MCO_N Mus musculus cDNA clone
 DEFINITION UI-M-AL1-ah1-c-07-0-UI 5', mRNA sequence.
 ACCESSION BE653345
 VERSION BE653345.1 GI:9979245
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclirognathi; Muridae; Murinae; Mus. 1 (bases 1 to 468)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene

discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477

COMMENT Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: mestr@nhi.nih.gov
 CDNA Library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements
 Seq primer: M13 Reverse.

FEATURES
 source Location/Qualifiers
 1..468
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-AL1-ah1-c-07-0-UI"
 /clone_lib="NIH_BMAP_MCO_N"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_MCO_N library is a normalized library constructed from mouse cortex. The tag is a string of 5 nucleotides present between the Not I site and the oligo-dT track. The library was constructed as described by Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996. Tissue provided by Ms. Annie Novakovich, Zivic-Miller Laboratories."

BASE COUNT 102 a 135 c 136 g 95 t

ORIGIN

Query Match 14.6%; Score 18; DB 10; Length 468;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 gagggctgactgactg 73
 |||
 Db 185 GAGGGCTGCACTGCTG 202

RESULT 18
 A0597171 487 bp DNA GSS 08-JUN-1999
 LOCUS HS-5240_A1_E12_SPE RPCI-11 Human Male BAC Library Homo sapiens
 DEFINITION genomic clone Plate-816 Col-23 Row-I, DNA sequence.
 ACCESSION A0597171
 VERSION A0597171.1 GI:5028383
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 487)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,V., Young,V., Zhao,S., Adams,M.D. and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome

JOURNAL Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)

MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618

Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm) or from Research Genetics (info@resgen.com). BAC end Web Server: http://www.htsc.washington.edu
 Plate: 816 row: 1 column: 23
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 487.

FEATURES

source

Location/Qualifiers

1..487

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="Plate=816 Col=23 Row=1"
 /clone_lib="RPCI-11 Human Male BAC Library"
 /sex="male"
 /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
 Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACe3.6 vector at EcoRI sites"

BASE COUNT

114 a 111 c 114 g 140 t 8 others

ORIGIN

Query Match 14.6%; Score 18; DB 13; Length 487;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 tcttcattatccaca 118
 |||||

Db 44 TCTTCATTATTCACA 27

RESULT 19

BH054207/c

LOCUS

DEFINITION

BH054207 533 bp DNA GSS 17-JUL-2001
 RPCI-24-337C12.TV RPCI-24 Mus musculus genomic clone RPCI-24-337C12
 , DNA sequence.

ACCESSION

BH054207

VERSION

BH054207.1

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

Zhaio.S., Nierman.W., Malek.J., Shatsman.S., Akınret.B., Levins.M., Tsegaye.G., Geer.K., Krol.M., Shvartsbeyn.A., Gebregiorgis.E., Russell.D., de Jong.P. and Fraser.C.M.

AUTHORS

TITLE

Mouse BAC End Sequences from Library RPCI-24

JOURNAL

Unpublished (1999)

COMMENT

Other_GSSs: RPCI-24-337C12.TV

CONTACT

Shaying Zhao

DEPARTMENT

Department of Eukaryotic Genomics

INSTITUTE

The Institute for Genomic Research

ADDRESS

9712 Medical Center Dr., Rockville, MD 20850, USA

TEL

301 838 0200

FAX

301 838 0208

EMAIL

szhao@tigr.org

CLONES

Clones are derived from the mouse BAC library RPCI-24. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://www.htsc.washington.edu/ordering_bac.htm) or from Research Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

PAGE

337 row: C column: 12

SEQ PRIMER

SP6

CLASS

BAC ends.

BASE COUNT

154 a 100 c 81 g 198 t

ORIGIN

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-24-337C12"
 /clone_lib="RPCI-24"
 /sex="Male"
 /cell_type="Spleen/Brain"
 /note="Vector: pTARBAC1; Site_1: BamHI; Site_2: BamHI;
 RPCI-24 Mouse BAC Library produced by Pieter de Jong. The library was cloned in the pTARBAC1 cloning vector at the BamHI sites using MboI partially digested male C57BL/6J DNA."

BASE COUNT

154 a 100 c 81 g 198 t

ORIGIN

Query Match 14.6%; Score 18; DB 13; Length 533;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 taagttcacatgtaag 91
 |||||

Db 266 TAACTTCACTAGG 249

RESULT 20

A2290583/c

LOCUS

DEFINITION

A2290583 574 bp DNA GSS 27-JUL-2000
 RPCI-23-129H21.TV RPCI-23 Mus musculus genomic clone RPCI-23-129H21
 , DNA sequence.

ACCESSION

A2290583

VERSION

A2290583.1

KEYWORDS

GSS.

SOURCE

house mouse.

ORGANISM

Mus musculus

REFERENCE

Zhaio.S., Nierman.W., Feldblyum.T., Malek.J., Shatsman.S., Akınret.B., Levins.M., McGann.S., Tsegaye.G., Geer.K., Krol.M., de Jong.P. and Fraser.C.M.

AUTHORS

TITLE

Mouse BAC End Sequences from Library RPCI-23

JOURNAL

Unpublished (1999)

COMMENT

Other_GSSs: RPCI-23-129H21.TV

CONTACT

Shaying Zhao

DEPARTMENT

Department of Eukaryotic Genomics

INSTITUTE

The Institute for Genomic Research

ADDRESS

9712 Medical Center Dr., Rockville, MD 20850, USA

TEL

301 838 0200

FAX

301 838 0208

EMAIL

szhao@tigr.org

CLONES

Clones are derived from the mouse BAC library RPCI-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Research Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html

PAGE

129 row: H column: 21

SEQ PRIMER

T7

CLASS

BAC ends.

FEATURES

source

Location/Qualifiers

1..574

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-129H21"
 /clone_lib="RPCI-23"
 /sex="female"
 /lab_host="DH10B"
 /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies).
 BASE COUNT 130 a 67 c 124 g 253 t
 ORIGIN

Query Match 14.6%; Score 18; DB 13; Length 574;
 Best Local Similarity 100.0%; Pred. No. 23;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 tcttcattattccaca 118
 |||
 Db 408 tcttcattattccaca 391

RESULT 21

AZ243191 595 bp DNA GSS 15-JUN-2000
 LOCUS RPCI-23-34A15.TV RPCI-23 Mus musculus genomic clone RPCI-23-34A15,
 DEFINITION DNA sequence.

ACCESSION AZ243191
 VERSION AZ243191.1 GI:8556298
 KEYWORDS GSS.

SOURCE

house mouse.
 Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sclurognathia; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 595)
 Zhao, S., Nieman, W., Feldblum, T., Malek, J., Shatsman, S., Akimret,
 'B., Levins, M., Mcgann, S., Tsegaye, G., Geer, R., Krol, M., de Jong, P.,
 and Fraser, C.M.

AUTHORS

Mouse BAC End Sequences from Library RPCI-23
 Unpublished (1999)

TITLE

Other GSS: RPCI-23-34A15.TV
 Contact: Shaying Zhao
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208

JOURNAL

Email: szhao@tigr.org

Clones are derived from the mouse BAC library RPCI-23. For BAC
 library availability, please contact Pieter de Jong
 (pieter@tigr.org), please contact Pieter de Jong
 (pieter@tigr.org) or from Resea ch Genetics (info@resgen.com). BAC end page:
 http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
 Plate: 34 row: A column: 15
 Seq primer: T7
 Class: BAC ends.

COMMENT

Location/Qualifiers
 1..595

/organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RPCI-23-34A15"
 /clone.lib="RPCI-23"
 /sex="Female"
 /lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site_1:
 EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
 brain genomic DNA was isolated and partially digested
 with a combination of EcoRI and EcoRI Methylase. Size
 selected DNA was cloned into the pBAC3.6 vector at the
 EcoRI sites. The ligation products were transformed into
 DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT

256 a 125 c 74 g 140 t

ORIGIN

Query Match 14.6%; Score 18; DB 13; Length 595;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 101 tcttcattattccaca 118
 |||
 Db 12 tcttcattattccaca 29

RESULT 22

BF815581/c 599 bp mRNA EST 13-JAN-2001
 LOCUS MR2-C10128-041200-009-c02 C10128 Homo sapiens cDNA, mRNA sequence.
 DEFINITION BF815581
 VERSION BF815581.1 GI:12150164
 KEYWORDS EST.

SOURCE

human.
 Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 599)
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,
 Brumstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,
 M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and
 Simpson, A.J.

AUTHORS

Shotgun sequencing of the human transcriptome with ORF expressed
 sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE

20202663

JOURNAL

CONTACT: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
 Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br

COMMENT

This sequence was derived from the FAPESP/LICR Human Cancer Genome
 Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?fl=MR2-C10128-
 041200-009-c02&t3=2000-12-04&t4=1)
 Seq primer: puc 18 forward
 High quality sequence stop: 599.

FEATURES

Location/Qualifiers
 1..599

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone.lib="C10128"
 /dev_stage="Adult"

source

/note="Organ: colon; Vector: puc18; Site_1: SmaI;
 Site_2: SmaI; A mini-library was made by cloning products
 derived from ORESTES PCR (U.S. Letters Patent application
 No. 196,716 - Ludwig Institute for Cancer Research)
 profiles into the puc 18 vector. Reverse transcription of
 tissue mRNA and cDNA amplification were performed under
 low stringency conditions."

BASE COUNT

156 a 137 c 151 g 132 t 3 others

ORIGIN

Query Match 14.6%; Score 18; DB 11; Length 599;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

BASE COUNT

256 a 125 c 74 g 140 t

ORIGIN

Query Match 14.6%; Score 18; DB 13; Length 595;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match

Best Local Similarity 100.0%; Pred. No. 24;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 23
 AZ396970 606 bp DNA GSS 03-OCT-2000
 LOCUS IM0161D13R Mouse 10kb plasmid uugc1m library Mus musculus genomic
 DEFINITION

clone UGCLM0161D13 R, DNA sequence.

ACCESSION A2396970
 VERSION A2396970.1 GI:10512042
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 606)
 Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamll, C., Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Kelly, M., Rose, M., Rose, R., Stokes, R., Tinney, A., von Niederhausen, A. and Wright, D., Weis, R.
 Mouse whole genome scaffolding with paired end reads from 10Kb plasmid inserts
 TITLE Unpublished (2000)
 JOURNAL COMMENT Contact: Robert B. Weiss
 University of Utah Genome Center
 Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLG, UT 84112, USA
 Tel: 801 585 5606
 Fax: 801 585 7177
 Email: ddunn@genetics.utah.edu
 Insert Length: 10000 Std Error: 0.00
 Plate: 0161 row: D column: 13
 Seq primer: CACACAGCAACAGCTATGACC
 Class: plasmid ends
 High quality sequence stop: 606.

FEATURES
 source
 Location/Qualifiers
 1..606
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UGCLM0161D13"
 /clone_lib="Mouse 10kb plasmid UGCLM library"
 /sex="Male"
 /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
 /note="Vector: pMD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource
 (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The sheared DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 147 a 95 c 107 g 257 t

ORIGIN

Query Match 14.6%; Score 18; DB 13; Length 606;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 ttatctcaatttatcc 115
 ||||||||||||||||
 Db 315 TTATCTCAATTATTC 332

RESULT 24
 BF208765/c 634 bp mRNA EST 06-NOV-2000
 LOCUS BF208765 634 bp mRNA EST 06-NOV-2000

601872235F1 NIH_MGC_53 Homo sapiens cDNA clone IMAGE:4093065 5', mRNA sequence.

ACCESSION BF208765
 VERSION BF208765.1 GI:11102351
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 634)
 NIH-MGC http://mgc.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL COMMENT Contact: Robert Strausberg, Ph.D.
 Email: c9apbs@email.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: Image.lnl.gov
 Plate: L1CM951 row: f column: 10
 High quality sequence stop: 489.

FEATURES
 source
 Location/Qualifiers
 1..634
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4093065"
 /clone_lib="NIH MGC 53"
 /tissue_type="carcinoma, cell line"
 /lab_host="DH10B (T1 phage-resistant)"
 /note="Organ: bladder; Vector: pDR-LIB (Clontech); Site_1: Sili (ggccgcctggcc); Site_2: Sili (ggccatggcc); Double-stranded cDNA was prepared from cell line RNA. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-ATTCTAGAGCCGCGCCGACATG-dT(30)-BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.55 kb (range 0.9-4.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

BASE COUNT 184 a 176 c 165 g 109 t

ORIGIN

Query Match 14.6%; Score 18; DB 11; Length 634;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 atgtaagtgatctat 101
 ||||||||||||||||
 Db 364 ATGTAAGTGATCTAT 347

RESULT 25
 AV326766 167 bp mRNA EST 11-NOV-1999
 LOCUS AV326766
 DEFINITION AV326766 RIKEN full-length enriched, adult male medulla oblongata
 Mus musculus cDNA clone 6330415D21 3', mRNA sequence.
 AV326766
 AV326766.1 GI:6366818

ACCESSION AV326766
 VERSION AV326766
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 167)
 Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukushima, Y., Hara, A., Hayatsu, N., Hirozane, T., Hori, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y.,

Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tateo, M., Tomioka, N., Tsunoda, Y., Matabiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.
RIKEN Mouse ESTs (Kono, H., et al. 1999)
Unpublished (1999)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
url: http://genome.gsc.riken.go.jp/
Sasaki, N., Izawa, M., Matabiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.

Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES

source

Location/Qualifiers
1. 167
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="6330415D21"
/clone_1lb="RIKEN full-length enriched, adult male medulla oblongata"
/sex="male"
/tissue_type="medulla oblongata"
/dev_stage="adult"
/lab_host="DH10B"
/note="Site 1: SalI; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTCGAGTTAATTAAATCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified Bluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

BASE COUNT

58 a 42 c 31 g 36 t

ORIGIN

Query Match 13.8%; Score 17; DB 10; Length 167;
Best Local Similarity 100.0%; Pred. No. 74;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 28 caaccctgatgacctaa 44
|||||
Db 25 CAACCTGATGACCTAA 41

RESULT 26

AA874072

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AA874072 211 bp mRNA EST 19-MAR-1998
vW88B10.r1 Striatum mouse skin (#937313) Mus musculus cDNA clone
IMAGE:1262011 5', mRNA sequence.
AA874072
AA874072.1 GI:2976761
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 211)
Marr, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Scheinberg, K., Stepien, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through INLND; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:664563
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 202.

FEATURES

source

Location/Qualifiers
1. 211
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:1262011"
/clone_1lb="Striatum mouse skin (#937313)"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: Bluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Whole skin from 11 week old C57BL/6 female mice. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5' CTCGACTTTTCTTTTCTTTT 3'."

BASE COUNT

42 a 25 c 47 g 97 t

ORIGIN

Query Match 13.8%; Score 17; DB 10; Length 211;
Best Local Similarity 100.0%; Pred. No. 75;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 62 tgcactgactgtaagt 78
|||||
Db 21 TGCACGTACTGTAAGT 37

RESULT 27

AV337137

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AV337137 249 bp mRNA EST 11-NOV-1999
AV337137 RIKEN full-length enriched, adult male medulla oblongata
Mus musculus cDNA clone 6332414J18 3', mRNA sequence.

Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 20.0 and subtraction to Rot = 479.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTCGAGTAAATTAATATCCCCCCCCCC 3']. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I."

BASE COUNT 87 a 88 c 39 g 63 t
ORIGIN

Query Match 13.8%; Score 17; DB 10; Length 277;
Best Local Similarity 100.0%; Pred. No. 77;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 caaccctgatgacctaa 44
|||||
Db 135 CAACCCGTGATGACCTAA 151

RESULT 29
AV277980 296 bp mRNA EST 05-NOV-1999
LOCUS AV277980 RIKEN full-length enriched, adult male testis (DH5a) Mus
DEFINITION musculus cDNA clone 4932703E19 3', mRNA sequence.
AV277980
VERSION AV277980.1 GI:6266017
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.

REFERENCE
AUTHORS Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 296)
Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F., Ishii, Y., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kojima, Y., Koya, S., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Ozawa, Y., Saito, H., Sano, M., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Takahashi, F., Tatenno, M., Tomimaga, N., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yasunishi, A., Yokota, T., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Kono, H., et al. 1999)
Unpublished (1999)

TITLE
JOURNAL
COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suenho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Sasaki, M., Izawa, M., Watahiki, M., Ozawa, K., Tanaka, T., Yoneda, Y., Matsura, S., Carninci, P., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)
Itoh, M., Katsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,

19-44 (1999)
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.
Location/Qualifiers

1. 296
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="4932703E19"
/clone_1bp="RIKEN full-length enriched, adult male testis (DH5a)"
/sex="male"
/tissue="testis"
/dev_stage="adult"
/lab_host="DH5a"

/note="Site 1: SalI. Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATCTCGAGTAAATTAATATCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI."

BASE COUNT 90 a 91 c 31 g 84 t
ORIGIN

Query Match 13.8%; Score 17; DB 10; Length 296;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 caaccctgatgacctaa 44
|||||
Db 154 CAACCCGTGATGACCTAA 170

RESULT 30
BB248570 296 bp mRNA EST 06-JUL-2000
LOCUS BB248570 RIKEN full-length enriched, 7 days neonate cerebellum Mus
DEFINITION musculus cDNA clone A730026M11 3', mRNA sequence.
BB248570
VERSION BB248570.1 GI:8941316
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.

REFERENCE
AUTHORS Mammalia: Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 296)
Kono, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T., Carninci, P., Endo, T., Fukuda, S., Fukunishi, Y., Hara, A., Hayatsu, N., Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M., Izawa, M., Kadota, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N., Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C., Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H., Okazaki, Y., Owa, C., Saito, H., Sakai, C., Sato, K., Shibata, K., Shibata, Y., Shigemoto, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A., Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watahiki, A., Watanabe, S., Yamamura, T., Yamane, I., Yano, R., Yasunishi, A., Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
RIKEN Mouse ESTs (Kono, H., et al.)
Unpublished (2000)

TITLE
JOURNAL

COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S., Sasaki
, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Thermostabilization and thermoactivation of thermostable enzymes by
trehalose and its application for the synthesis of full length
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
Itoh, M., Kitsunai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki
, Y. and Hayashizaki, Y.
Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)
Please visit our web site (http://genome.riken.go.jp) for
further details.

FEATURES
SOURCE
Location/Qualifiers
1..296
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="A730026M11"
/clone_lib="RIKEN full-length enriched, 7 days neonate
cerebellum"
/tissue_type="cerebellum"
/dev_stage="7 days neonate"
/lab_host="DH10B"
/note="Site 1: Salt; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 20.0 and subtraction to Rot = 459.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGATCCAGAGTATTAATTAATCCCCCCCCCCC
3']. cDNA was cleaved with XhoI and BamHI. Vector: a
modified pBluescript KS(+) after bulk excision from Lambda
FLC I."

BASE COUNT 97 a 69 c 42 g 88 t
ORIGIN

Query Match 13.8%; Score 17; DB 10; Length 296;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 28 caaccctgatgacctaa 44
|||||
Db 156 CAACCTGATGACTTA 172

RESULT 31
LOCUS AI466360 301 bp mRNA EST 09-MAR-1999
DEFINITION v888b10.y1 StrataGene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:1262011 5', mRNA sequence.
ACCESSION AI466360
VERSION AI466360.1 GI:4314979
KEYWORDS EST.
SOURCE house mouse.

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person
, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter
, E., Korn, S., Shin, T., Jackson, J., Cardenas, M., McCann, R.,
Waterston, R. and Wilson, R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
COMMENT Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (Info@image.lnl.gov) for further information.
MGI:664563
This read is a RESEQUENCE of a previously sequenced mouse clone
This read has been verified (found to hit its original self in the
correct orientation)
Seq primer: -40RP from Gibco
High quality sequence stop: 283.

FEATURES
SOURCE
Location/Qualifiers
1..301
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:1262011"
/clone_lib="Stratagene mouse skin (#937313)"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site 1: EcoRI
; Site 2: XhoI; Cloned unidirectionally. Primer: Oligo
dT. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGACGACGAC 3' -3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' "

BASE COUNT 62 a 41 c 66 g 130 t
ORIGIN

Query Match 13.8%; Score 17; DB 10; Length 301;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 62 tgcaactgactgtaagt 78
|||||
Db 71 TGCACCTGACTGTAAGT 87

RESULT 32
LOCUS B1188737/c 316 bp mRNA EST 10-JUL-2001
DEFINITION d2904fs.r1 Fusarium sporotrichioides Tr1 10 overexpressed cDNA
library Fusarium sporotrichioides cDNA clone d2904fs 5', mRNA
sequence.
ACCESSION B1188737
VERSION B1188737.1 GI:14662416
KEYWORDS EST.
SOURCE Fusarium sporotrichioides.
ORGANISM Fusarium sporotrichioides.
REFERENCE Hypocretales; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
1 (bases 1 to 316)
AUTHORS Ren, Q., Tag, A., Pehlows, A., Lal, H., Kupfer, C., Peterson, A., Beremand
, M. and Roe, B.
TITLE Analysis of a Fusarium sporotrichioides EST database
JOURNAL Unpublished (2001)

COMMENT
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
Contact Dr. Marian Beremand regarding clone availability. Included
is the best homolog from a blastx search of Genbank nr 04-09-01
[Arabidopsis thaliana]
Seq primer: T3
High quality sequence stop: 136.
Location/Qualifiers
1. 316
/organism="Fusarium sporotrichioides"
/strain="Tri 10"
/db_xref="taxon:5514"
/clone="q2g04fs"
/clone_lib="Fusarium sporotrichioides Tri 10 overexpressed
cDNA library"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
; 3' end of cDNA cloned into XhoI site of pBluescript"
BASE COUNT 82 a 102 c 63 g 69 t
ORIGIN

Query Match 13.8%; Score 17; DB 11; Length 316;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 53 tgcgcagcgcgcactga 69
|||||
Db 64 TCGCAGCGCTGCACTGA 48

RESULT 33
BI266425 329 bp mRNA EST 18-JUL-2001
LOCUS
DEFINITION NF097C04IN.F1034 Insect herbivory Medicago truncatula cDNA clone
ACCESSION BI266425
VERSION BI266425.1 GI:14870518
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
1 (bases 1 to 329)
Korth, K., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores
H.R., Inman, J.T., Weller, J.W. and May, G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula insect herbivory library
Unpublished (2000)
JOURNAL
COMMENT
Contact: Korth K
Dept. of Plant Pathology
University of Arkansas
217 Plant Science Building, Fayetteville, AR 72701, USA
Tel: 501 575 5191
Fax: 501 575 7601
Email: kthorne@comp.uark.edu
Insert Length: 329 Std Error: 0.00
Plate: 097 row: C column: 04
Seq primer: TCACACAGGAGAACAGCTATGAC.
FEATURES
source
1. 329
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF097C04IN"
/clone_lib="insect herbivory"

/tissue_type="local and systemic leaves"
/dev_stage="mature"
/note="Vector: Lambda Zap. library was produced from fully
expanded M. truncatula leaves of plants fed upon by
spodoptera exigua (beet armyworm) for 24 hours. Systemic
(undamaged leaves from injured plants) and wounded leaves
were harvested and pooled."

BASE COUNT 111 a 62 c 50 g 104 t 2 others
ORIGIN

Query Match 13.8%; Score 17; DB 11; Length 329;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 101 tcttcaattatccac 117
|||||
Db 42 TCTTCATTATTCAC 58

RESULT 34
AA244200/c 334 bp mRNA EST 20-AUG-1997
LOCUS
DEFINITION nc06c06.s1 NCI-CGAP_Prl Homo sapiens cDNA clone IMAGE:1007338
similar to contains Alu repetitive element; contains element MER35
repetitive element ;, mRNA sequence.
ACCESSION AA244200
VERSION AA244200.1 GI:1874959
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 334)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index.
Unpublished (1997)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: W. Marston Linehan, M.D., Rodrigo Chuagui, M.D.
Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -41ml3 fwd. ET from Amersham
High quality sequence stop: 293.
Location/Qualifiers
1. 334
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1007338"
/clone_lib="NCI-CGAP_Prl"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: PAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from 5,000-10
,000 microdissected, histologically normal prostate
epithelial cells. Double-stranded cDNA was ligated to
EcoRI adaptors, 5 cycles of PCR applied to the cDNA with
an adaptor-specific primer, and the resulting PCR product
subcloned into PAMP10 by the upc-cloning method (Life
Technologies). Average insert size is 600 bp. NOTE: Not
directionally cloned. This library was constructed by
David Krizman."

BASE COUNT 111 a 66 c 82 g 75 t
ORIGIN

Query Match 13.8%; Score 17; DB 10; Length 334;
 Best Local Similarity 100.0%; Pred. No. 78;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 99 tatcttaattattcc 115
 |||
 Db 33 TATCTCAATTATTC 17

RESULT 35
 C70635/c 360 bp mRNA EST 29-SEP-1997
 LOCUS C70635 yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
 DEFINITION clone yK411b1 5', mRNA sequence.
 ACCESSION C70635
 VERSION C70635.1 GI:2441160
 KEYWORDS EST.
 SOURCE Caenorhabditis elegans.
 ORGANISM Caenorhabditis elegans.
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea
 ; Rhabditidae; Peloderiinae; Caenorhabditis.
 REFERENCE 1 (bases 1 to 360)
 AUTHORS Kohara,Y., Mochizashi,T., Tabara,H., Watanabe,H., Sugimoto,A., Sano
 TITLE Expression map of the C.elegans genome
 JOURNAL Unpublished (1996)
 COMMENT Contact: Yuji Kohara
 Genome Biology Lab.
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 81-559-81-6854
 Fax: 81-559-81-6855
 Email: ykohara@lab.nig.ac.jp.

FEATURES
 source location/Qualifiers
 1..360
 /organism="Caenorhabditis elegans"
 /strain="CB1489 him-8(e1489)"
 /db_xref="taxon:6239"
 /clone="YK411b1"
 /clone.lib="yuji Kohara unpublished cDNA"
 /sex="hermaphrodite, male"
 /tissue_type="whole animal"
 /dev_stage="varied"

BASE COUNT 102 a 63 c 79 g 106 t 10 others
 ORIGIN

Query Match 13.8%; Score 17; DB 11; Length 360;
 Best Local Similarity 100.0%; Pred. No. 79;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 105 caattattcacacaa 121
 |||
 Db 287 CAATTATTCACACAA 271

RESULT 36
 B1266362 366 bp mRNA EST 18-JUL-2001
 LOCUS B1266362
 DEFINITION NF092D091NF1078 Insect herbivory Medicago truncatula cDNA clone
 NF092D091N 5', mRNA sequence.
 ACCESSION B1266362
 VERSION B1266362.1 GI:14870403
 KEYWORDS EST.
 SOURCE barrel medic.
 ORGANISM Medicago truncatula
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
 Medicago.
 REFERENCE 1 (bases 1 to 366)
 AUTHORS Korth,K., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J., Flores

TITLE 'H.R., Inman,J.T., Weller,J.W. and May,G.D.
 JOURNAL Expressed Sequence Tags from the Samuel Roberts Noble Foundation
 COMMENT Medicago truncatula insect herbivory library
 Unpublished (2000)
 Contact: Korth K
 Dept. of Plant Pathology
 University of Arkansas
 217 Plant Science Building, Fayetteville, AR 72701, USA
 Tel: 501 575 5191
 Fax: 501 575 7601
 Email: Kkorth@comp.uark.edu
 Insert Length: 366 Std Error: 0.00
 Plate: 092 row: D column: 09
 Seq primer: TCACACGACGAACACCTATGAC.
 Location/Qualifiers
 1..366
 /organism="Medicago truncatula"
 /db_xref="taxon:3880"
 /clone="NF092D091N"
 /clone.lib="Insect herbivory"
 /tissue_type="local and systemic leaves"
 /dev_stage="mature"
 /note="Vector: Lambda Zap; Library was produced from fully
 expanded M. truncatula leaves of plants fed upon by
 Spodoptera exigua (beet armyworm) for 24 hours. Systemic
 (undamaged leaves from injured plants) and wounded leaves
 were harvested and pooled."

BASE COUNT 115 a 69 c 52 g 128 t 2 others
 ORIGIN

Query Match 13.8%; Score 17; DB 11; Length 366;
 Best Local Similarity 100.0%; Pred. No. 79;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 101 tcttcaattattccac 117
 |||
 Db 104 TCTTCAATTATTCAC 120

RESULT 37
 AW457639 450 bp mRNA EST 24-FEB-2000
 LOCUS AW457639
 DEFINITION UI-M-BH3-ari-a-04-0-01.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
 UI-M-BH3-ari-a-04-0-01 3', mRNA sequence.
 ACCESSION AW457639
 VERSION AW457639.1 GI:7027896
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 450)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: mestr@nmi.nih.gov

The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to identify it as a clone from the
 normalized corpus striatum library cDNA Library Preparation: M.B.
 Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
 clones from RESEARCH GENETICS. It should be noted that Bento Soares

is generating a small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLY-A-Tes.

FEATURES

source

Location/Qualifiers
1. .456
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BH3-art-a-04-0-01"
/clone_lib="NIH-BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p773D-Pac (Pharmacia) with a modified polylinker. Site_1: Not I; Site_2: Eco RI. The NIH-BMAP_M_S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hippocampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH-BMAP_M_S4, NIH-BMAP_M_S3.3, NIH-BMAP_M_S3.2, NIH-BMAP_M_S3.1, NIH-BMAP_M_S2, NIH-BMAP_M_S1. The subtracted library (NIH-BMAP_M_S4) was constructed as follows: PCR amplified cDNA inserts from NIH-BMAP_M_S3.3, NIH-BMAP_M_S3.2, and NIH-BMAP_M_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH-BMAP_M_S3.3, NIH-BMAP_M_S3.2, and NIH-BMAP_M_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the NIH-BMAP_M_S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=NIH-BMAP_M_S4
TAG_TISSUE=corpus-striatum
TAG_SEQ=ACGCGC"

BASE COUNT 110 a 127 c 109 g 104 t
ORIGIN

Query Match 13.8%; Score 17; DB 10; Length 450;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 gactgtaagtccaca 84
|||||
Db 377 GACTGTAGTTTCACA 393

RESULT 38

LOCUS BE649642 456 bp mRNA EST 20-DEC-2000
DEFINITION NF082D01EC1F1012 Elicited cell culture Medicago truncatula cDNA
ACCESSION BE649642
VERSION BE649642.1 GI:11914772
KEYWORDS EST
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.
1 (bases 1 to 456)
REFERENCE
Torres-Jerez, I., Scott, A.D., Harris, A.R., Gonzales, R.A., Bell, C.J., Flores, H.R., Inman, J.T., Weller, J.W. and May, G.D.

TITLE
JOURNAL
COMMENT
Expressed Sequence Tags from the Samuel Roberts Noble Foundation - Center for Medicago Genomics Research
Unpublished (2000)
Contact: Dixon RA
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7302
Fax: 580 221 7380
Email: radixon@noble.org
Insert Length: 456 Std Error: 0.00
Plate: 082 row: D column: 01
Seq primer: TCACACGAGAACACGCTATGAC.

FEATURES

source

Location/Qualifiers
1. .456
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF082D01EC"
/clone_lib="Elicited cell culture"
/tissue_type="Cell cultures derived from root tissues"
/dev_stage="Cell suspensions were subcultured every 14 days. Cells were induced six days after subculture"
/note="Vector: Lambda Zap; Cells were induced with yeast cell wall extracts equivalent to 50ug/ml glucose in the final concentration. Samples were taken at 0.5', 1', 12 and 24 hours after induction. Equal amounts of RNA from each time point were pooled and used for mRNA isolation."

BASE COUNT 146 a 86 c 66 g 158 t
ORIGIN

Query Match 13.8%; Score 17; DB 11; Length 456;
Best Local Similarity 100.0%; Pred. No. 80;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 101 tcttcaattatccac 117
|||||
Db 120 TCTTCATTATTCAC 136

RESULT 39

LOCUS BE989827 489 bp mRNA EST 05-OCT-2000
DEFINITION UI-M-BZ1-bft-e-20-0-UI s1 NIH-BMAP_MH12_S1 Mus musculus cDNA clone
ACCESSION BE989827
VERSION BE989827.1 GI:10667610
KEYWORDS EST
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 489)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene discovery
JOURNAL
MEDLINE
COMMENT
Genome Res. 6 (9), 791-806 (1996)
9704477
Contact: Chin, H
National Institute of Mental Health
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD 20892-9643, USA
Tel: 301 443 1706
Fax: 301 443 9890
Email: mestr@mail.nih.gov
The sequence contained an oligo-dr track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dr track served to verify it as a clone from the hippocampus tissue cDNA library Preparation: M.B. Soares Lab Clone distribution: Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It should be noted that Bento Soares is generating a

small number of additional specialized non-redundant arrays of BMAP cDNAs whose availability will be considered under appropriate and limited collaborative arrangements
Seq primer: M13 Forward
POLY-A=yes

FEATURES

Source

Location/Qualifiers
1. 489
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UI-M-BZ1-Dft-e-20-0-UI"
/clone_1ib="NIH_BMAP_MH12_S1"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_MH12_S1 library is a subtracted library derived from NIH_BMAP_MH12. NIH_BMAP_MH12 is a library derived from mouse hippocampus tissue. For a detailed description of the library from which this clone was derived, please visit our web site at brainest.eng.uiowa.edu.
TAG_LIB=NIH_BMAP_MH12_S1
TAG_TISSUE=hippocampus
TAG_SEQ=TACTC"

BASE COUNT 118 a 142 c 123 g 106 t
ORIGIN

Query Match 13.8%; Score 17; DB 11; Length 489;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 68 gactgtaagtttcaca 84
|||||
Db 356 GACTGTAGTTTCA 372

RESULT 40
AM394535 519 bp mRNA EST 07-FEB-2000
LOCUS sh32c05.y1 Gm-cl017 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-cl017-3393 5' similar to TR:083191 083191 CONSERVED HYPOTHETICAL
PROTEIN. ; mRNA sequence.
ACCESSION AM394535
VERSION AM394535.1 GI:6912947
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 519)
Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
,A., Bolla,B., Marr,M., Hiller,L., Kucaba,T., Martin,J., Beck,C.,
Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R./Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
Info@genomesystems.com web site: www.genomesystems.com

Seq primer: -40RP from Gibco
High quality sequence stop: 426.
Location/Qualifiers
1. 519
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl017-3393"
/clone_1ib="Gm-cl017"
/tissue_type="vegetable buds of field grown plants"
/lab_host="Xl10-Gold"
/note="Vector: p Bluescript II XR; Site_1: EcoRI; Site_2:
XhoI. This cDNA library was constructed from mRNA isolated
from vegetable buds of field grown plants. The cDNA
library was prepared using the Stratagene p Bluescript II
XR library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dT) sequence with a XhoI restriction site. EcoRI adapters
were ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
p Bluescript vector. The ligated cDNA fragments were
transformed into Xl10-Gold host cells. This library was
constructed by Dr. Randy Shoemaker and Dr. John
Erpelting."

BASE COUNT 143 a 91 c 120 g 164 t 1 others
ORIGIN

Query Match 13.8%; Score 17; DB 10; Length 519;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 95 tactatcttcaattta 111
|||||
Db 488 TACTATCTTCAATTTA 504

RESULT 41
A2626769 526 bp DNA GSS 13-DEC-2000
LOCUS A2626769/c
DEFINITION M0467008F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0467008 F, DNA sequence.
ACCESSION A2626769
VERSION A2626769.1 GI:11748959
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 526)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamill,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
Plasmid Inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0467 row: 0 column: 08
Seq primer: CGTTGTAAACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 526.
Location/Qualifiers
1. 526
/organism="Mus musculus"

TITLE
JOURNAL
COMMENT

FEATURES
Source


```

/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_1lb="UGC1M0467008"
/clone_1lb="Mouse 10kb plasmid UGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance."
BASE COUNT      215 a      97 c      92 g      122 t
ORIGIN

```

```

Query Match      13.8%; Score 17; DB 13; Length 526;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 100 attcattatcaca 116
      |||||||
Db 179 ATCTCAATTATTCACA 163

```

```

RESULT 42
BG048538      541 bp      mRNA      EST      25-JAN-2001
LOCUS
DEFINITION
OVL_14_H02_g2_A002 Ovary 1 (OVL) Sorghum bicolor cDNA, mRNA
sequence.
ACCESSION
BG048538
VERSION
BG048538.1 GI:12499394
KEYWORDS
EST.
SOURCE
sorghum.
ORGANISM
Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Sorghum.
1 (bases 1 to 541)
Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
,L.H.
An EST database from Sorghum: ovaries of varying immature stages
Unpublished (2000)
Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmp@prattuga.edu
Sequences have been trimmed to exclude PolyA, vector and regions
below phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTmix
High quality sequence start: 30
High quality sequence stop: 541
POLYA-NO.
Location/Qualifiers
1. 541
/organism="Sorghum bicolor"

```

```

FEATURES
source
1. 541
/organism="Sorghum bicolor"

```

```

/db_xref="taxon:4558"
/clone_1lb="Ovary 1 (OVL)"
/notes="Organ: Mix of ovaries of varying immature stages
from 8-week-old plants; Vector: pBluescript II from lambda
Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced were prepared by mass excision."
BASE COUNT      139 a      111 c      141 g      150 t
ORIGIN

```

```

Query Match      13.8%; Score 17; DB 11; Length 541;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 26 tacacctgatgacct 42
      |||||||
Db 74 TACAACTGTGATGACCT 90

```

```

RESULT 43
AZ961481      544 bp      DNA      GSS      27-APR-2001
LOCUS
DEFINITION
2M0229L22R Mouse 10kb plasmid UGC2M library Mus musculus genomic
clone UGC2M0229L22 R; DNA sequence.
ACCESSION
AZ961481
VERSION
AZ961481.1 GI:13832708
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognath; Muridae; Murinae; Mus.
1 (bases 1 to 544)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingy,A., von Niederhausen,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0229 row: 1 column: 22
Seq primer: CACACAGGAACACGCTATGACC
Class: plasmid ends
High quality sequence stop: 544.
Location/Qualifiers
1. 544
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_1lb="UGC2M0229L22"
/clone_1lb="Mouse 10kb plasmid UGC2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel

```

```

FEATURES
source
1. 544
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_1lb="UGC2M0229L22"
/clone_1lb="Mouse 10kb plasmid UGC2M library"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/notes="Vector: PWD42nv; Purified genomic DNA from M. musculus C57BL/6J (female) was obtained from the Jackson Laboratory Mouse DNA Resource (http://www.jax.org/resources/documents/dnares/). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel

```

electrophoresis. Vector DNA was prepared from a derivative of pMD42 (g114732114|9b|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

BASE COUNT 166 a 139 c 101 g 138 t
ORIGIN

Query Match 13.8%; Score 17; DB 13; Length 544;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 69 actgtaatgtaacat 85
|||||
Db 426 ACTGTAATGTCACAT 442

RESULT 44
A2561989

LOCUS RPT-23-21704.TV RPT-23 Mus musculus genomic clone RPT-23-21704,
DEFINITION DNA sequence.
ACCESSION A2561989
VERSION A2561989.1 GI:11241809
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 555)
AUTHORS Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S., Akimret,
B., Levins,M., McGann,S., Tsengaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPT-23
Unpublished (1999)
Contact: Shaying Zhao
Other GSSs: RPT-23-21704.TV
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208

TITLE
JOURNAL
COMMENT
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPT-23. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.bufileo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.bufileo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/tldb/bac-ends/mouse/bac_end_intro.html
Plate: 217 row: 0 column: 4
Seq primer: SP6
Class: BAC ends.

FEATURES
source
Location/Qualifiers
1..555

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPT-23-21704"
/clone_lib="RPT-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pACe3.6; Site:1;
ECORI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pACe3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT

193 a 108 c 116 g 138 t

ORIGIN

Query Match 13.8%; Score 17; DB 13; Length 555;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 62 tgcactgactgtaagt 78
|||||
Db 535 TGCACGTACTGTGTAAGT 551

RESULT 45
A2149056

LOCUS A2149056 557 bp DNA GSS 28-AUG-2000
DEFINITION SP.0046.A1.B09.T7 Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone plate=46 Col-17 Row-C, DNA sequence.
ACCESSION A2149056
VERSION A2149056.1 GI:8300957
KEYWORDS GSS.
SOURCE Strongylocentrotus purpuratus.
ORGANISM Strongylocentrotus purpuratus

REFERENCE 1 (bases 1 to 557)
AUTHORS Cameron,R.A., Mahairs,C., Rast,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray, G.A., Ettensohn,C.A., Lehrach,H., Bittlen,R.J., Davidson,E.H. and Hood,L.

TITLE
JOURNAL
MEDLINE
COMMENT
A sea urchin genome project: Sequence scan, virtual map, and additional resources
Proc. Natl. Acad. Sci. U. S. A. 97 (17), 9514-9518 (2000)
Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
Pasadena California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 46 row: C column: 17
Seq primer: T7
Class: BAC ends
High quality sequence strop: 557.

FEATURES
source
Location/Qualifiers
1..557

/organism="Strongylocentrotus purpuratus"
/db_xref="taxon:7668"
/clone="Plate=46 Col-17 Row=C"
/clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BACe3.6; BAC Clones in E-Coli DH10B"

BASE COUNT 177 a 111 c 114 g 152 t 3 others
ORIGIN

Query Match 13.8%; Score 17; DB 13; Length 557;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 tattgtaatgtaagat 21
|||||
Db 455 TATTGTAATGTGACAT 471

Search completed: March 19, 2002, 10:40:41
Job time: 3945 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2002, 09:58:46 ; Search time 1344.34 Seconds
(without alignments)
1509.405 Million cell updates/sec

Title: US-09-684-016-48411

Perfect score: 123

Sequence: 1 caggtatgtgtaaatgtgaga.....tcatttattccacacacaca 123

Scoring table: OLIGO_NUC

Searched: 1472140 segs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_hgo_hum:*
31: em_hgo_inv:*
32: em_hgo_rod:*
33: em_hcg_hum:*
34: em_hcg_inv:*
35: em_hcg_rod:*
36: em_hcg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	21	17.1	69734	2	AC010953	AC010953 Homo sapi
2	21	17.1	178035	2	AC068680	AC068680 Homo sapi
3	21	17.1	183079	2	AC007600	AC007600 Homo sapi
4	21	17.1	194198	2	AC013298	AC013298 Homo sapi
5	20	16.3	50974	9	AP000958	AP000958 Homo sapi
6	20	16.3	155304	2	AC024713	AC024713 Homo sapi
7	20	16.3	158927	2	CNS07EFA	CNS07EFA Homo sapi
8	20	16.3	171345	9	CNS01RIT	AL151361 Homo sapi
9	20	16.3	182340	2	AC024417	AC024417 Homo sapi
10	20	16.3	340000	9	AP001675	AP001675 Homo sapi
11	19	15.4	27710	9	U41746	U41746 Caenorhabd
12	19	15.4	130057	9	AL357912	AL357912 Human DNA
13	19	15.4	180911	2	AC027498	AC027498 Homo sapi
14	18	14.6	11018	6	AX073208	AX073208 Sequence
15	18	14.6	11018	1	AE002395	AE002395 Neisseria
16	18	14.6	33803	3	CELC34H3	AC066614 Caenorhab
17	18	14.6	34128	9	AC022128	AC022128 Homo sapi
18	18	14.6	57000	9	AC078862	AC078862 Homo sapi
19	18	14.6	80374	8	T8K14	AC007202 Arabidops
20	18	14.6	117018	9	AC008790	AC008790 Homo sapi
21	18	14.6	140952	8	AP003046	AP003046 Oryza sat
22	18	14.6	144536	9	AL161454	AL161454 Human DNA
23	18	14.6	146571	2	AL359815	AL359815 Homo sapi
24	18	14.6	148498	2	AC034225	AC034225 Homo sapi
25	18	14.6	153405	2	AL357252	AL357252 Homo sapi
26	18	14.6	156795	2	AC023537	AC023537 Homo sapi
27	18	14.6	165858	2	AC024315	AC024315 Homo sapi
28	18	14.6	183496	2	AC009776	AC009776 Homo sapi
29	18	14.6	190663	2	AC068581	AC068581 Homo sapi
30	18	14.6	192430	2	AC079542	AC079542 Mus muscu
31	18	14.6	198605	2	AL357568	AL357568 Homo sapi
32	18	14.6	218073	2	AC025501	AC025501 Mus muscu
33	18	14.6	226060	2	AC079583	AC079583 Mus muscu
34	18	14.6	349680	6	AX044029	AX044029 Sequence
35	17	13.8	518	6	AX094489	AX094489 Sequence
36	17	13.8	2195	20	AY004261	AY004261 Heterocap
37	17	13.8	34544	3	CELY58G8A	AC006808 Caenorhab
38	17	13.8	38225	9	AL591603	AL591603 Human DNA
39	17	13.8	44751	8	SPC116C4	AL031535 S. pombe c
40	17	13.8	50821	8	AC020579	AC020579 Arabidops
41	17	13.8	59935	2	AC018376	AC018376 Homo sapi
42	17	13.8	63604	8	AP000381	AP000381 Arabidops
43	17	13.8	65857	2	AC018413	AC018413 Homo sapi
44	17	13.8	65998	2	AC023633	AC023633 Homo sapi
45	17	13.8	70677	2	AC027574	AC027574 Homo sapi

ALIGNMENTS

RESULT	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
1.	AC010953	Homo sapiens clone 5_H_9, LOW-PASS SEQUENCE SAMPLING.	AC010953	GI:5931409	HTG: HTGS_PHASE0.	human.	Homo sapiens	28-SEP-1999	
	AC010953	69734 bp DNA					Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
	AC010953	1 (bases 1 to 69734)					Birren,B., Linton,L., Nussbaum,C. and Lander,E.		
	AC010953	2 (bases 1 to 69734)					Birren,B., Linton,L., Nussbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barina,N., Becker,I., Boguslavsky,L., Boukhalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,		

TITLE	Direct Submission
JOURNAL	Submitted (28-SEP-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT	All repeats were identified using RepeatMasker: Smit, A.F.A. &

Green, P. (1996-1997)

TITLE
JOURNAL

JOURNAL

COMMENT

Ferreira, P., FiltzHugh, W., Forrest, C., Funke, R., Gage, D.,
 Galagan, J., Gairdys, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
 Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
 Lechoczky, J., Lien, C., Locke, K., Macdonald, P., Margulis, N.,
 McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
 Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
 Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
 Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Testa, S., Tittrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
 Wymen, D., Ye, W. J., Zimmer, A. and Zody, M.
 Direct Submission
 Submitted (28-SEP-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker: Smit, A.F.A. &
 Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>.
 * NOTE: This record contains 78 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved

*	1	933:	config of 933 bp in length
*			gap of unknown length
*	934	1798:	config of 865 bp in length
*			gap of unknown length
*	1799	2683:	config of 885 bp in length
*			gap of unknown length
*	2684	3569:	config of 886 bp in length
*			gap of unknown length
*	3570	4443:	config of 874 bp in length
*			gap of unknown length
*	4444	5297:	config of 854 bp in length
*			gap of unknown length
*	5298	6183:	config of 886 bp in length
*			gap of unknown length
*	6184	7057:	config of 874 bp in length
*			gap of unknown length
*	7058	7989:	config of 932 bp in length
*			gap of unknown length
*	7990	8859:	config of 870 bp in length
*			gap of unknown length
*	8860	9721:	config of 862 bp in length
*			gap of unknown length
*	9722	10593:	config of 872 bp in length
*			gap of unknown length
*	10594	11472:	config of 879 bp in length
*			gap of unknown length
*	11473	12381:	config of 909 bp in length
*			gap of unknown length
*	12382	13275:	config of 894 bp in length
*			gap of unknown length
*	13276	14178:	config of 903 bp in length
*			gap of unknown length
*	14179	15072:	config of 894 bp in length
*			gap of unknown length
*	15073	15954:	config of 882 bp in length
*			gap of unknown length
*	15955	16816:	config of 862 bp in length
*			gap of unknown length
*	16817	17750:	config of 934 bp in length
*			gap of unknown length
*	17751	18613:	config of 863 bp in length
*			gap of unknown length
*	18614	19490:	config of 877 bp in length
*			gap of unknown length
*	19491	20372:	config of 882 bp in length
*			gap of unknown length

20373	21258:	contig of 886 bp in length
21259	22199:	contig of 941 bp in length
22200	23107:	gap of unknown length
23108	24014:	gap of unknown length
24015	24893:	contig of 907 bp in length
24894	25753:	gap of unknown length
25754	26631:	contig of 878 bp in length
26632	27505:	gap of unknown length
27506	28396:	gap of unknown length
28397	29283:	gap of unknown length
29284	30172:	contig of 889 bp in length
30173	31120:	gap of unknown length
31121	32109:	contig of 948 bp in length
32110	33003:	gap of unknown length
33004	33886:	contig of 883 bp in length
33887	34768:	gap of unknown length
34769	35647:	contig of 882 bp in length
35648	36542:	gap of unknown length
36543	37411:	contig of 879 bp in length
37412	38287:	gap of unknown length
38288	39172:	contig of 876 bp in length
39173	40064:	gap of unknown length
40065	40944:	contig of 885 bp in length
40945	41799:	gap of unknown length
41800	42711:	contig of 855 bp in length
42712	43589:	gap of unknown length
43590	44511:	contig of 912 bp in length
44512	45377:	gap of unknown length
45378	46252:	contig of 878 bp in length
46253	47126:	gap of unknown length
47127	48022:	contig of 874 bp in length
48023	48921:	gap of unknown length
48922	49794:	contig of 899 bp in length
49795	50651:	gap of unknown length
50652	51530:	contig of 873 bp in length
51531	52415:	gap of unknown length
52416	53304:	contig of 885 bp in length
		gap of unknown length
		contig of 889 bp in length

```

*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      53305      54179: contig of 875 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      54180      55086: contig of 907 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      55087      55996: contig of 910 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      55997      56897: contig of 901 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      56898      57842: contig of 945 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      57843      58797: contig of 955 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      58798      59688: contig of 891 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      59689      60549: contig of 861 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      60550      61436: contig of 887 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      61437      62316: contig of 880 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      62317      63198: contig of 882 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      63199      64114: contig of 916 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      64115      64983: contig of 869 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      64984      66019: contig of 1036 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      66020      66930: contig of 911 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      66931      67922: contig of 992 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      67923      68838: contig of 916 bp in length      gap of unknown length
*      *      *      *      *      *      *      *      *      *      *      *      *      *      *      *
*      68839      69734: contig of 896 bp in length.

```

Query Match 17.1%; Score 21; DB 2; Length 69734;

Best Local Similarity 100.0%; Pred. No. 0.35; Mismatches 0; Indels 0; Gaps 0;

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 102 cttcaattttccacacacac 122

Db 15008 CTTCAATTATTCACACACAC 15028

RESULT 2

LOCUS AC068680 178035 bp DNA HTG 04-JUN-2000
DEFINITION Homo sapiens clone RP11-22D13, WORKING DRAFT SEQUENCE, 11 unordered
pieces.

AC068680.3 GI:8247854
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM

REFERENCE 1 (bases 1 to 178035)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Mamaliya, E., Euterli, P., Primate, C., Catarrhini, H., Homindae, Homo.
TITLE Homo sapiens, clone RP11-22D13
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 178035)
TITLE Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Mamaliya, E., Euterli, P., Primate, C., Catarrhini, H., Homindae, Homo.
JOURNAL Unpublished

REFERENCE 1 (bases 1 to 178035)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Mamaliya, E., Euterli, P., Primate, C., Catarrhini, H., Homindae, Homo.
TITLE Homo sapiens, clone RP11-22D13
JOURNAL Unpublished

FEATURES

source

Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPeckers, R.,
Meldrum, J., Menius, L., Milnova, T., Miranda, C., Mieng, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tefaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Submitted (06-MAY-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 4, 2000 this sequence version replaced gi:7798779.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: LA102
Center clone name: 22_D_13
----- Summary Statistics
Sequencing vector: M13; M7815, 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 169315 bases at least Q40
Consensus quality: 173916 bases at least Q30
Consensus quality: 175957 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 177035; sum-of-contigs
Quality coverage: 4.7 in Q20 bases; agarose-fp
Quality coverage: 4.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1212: contig of 1212 bp in length
* 1213 1312: gap of 100 bp
* 1313 5818: contig of 4506 bp in length
* 5819 5918: gap of 100 bp
* 5919 9462: contig of 3544 bp in length
* 9463 9562: gap of 100 bp
* 9563 15211: contig of 5649 bp in length
* 15212 15311: gap of 100 bp
* 15312 28255: contig of 12944 bp in length
* 28256 28355: gap of 100 bp
* 28356 44646: contig of 16291 bp in length
* 44647 44746: gap of 100 bp
* 44747 63409: contig of 18663 bp in length
* 63410 63509: gap of 100 bp
* 63510 90494: contig of 26985 bp in length
* 90495 90594: gap of 100 bp
* 90595 118529: contig of 27935 bp in length
* 118530 118629: gap of 100 bp
* 118630 145844: contig of 27215 bp in length
* 145845 145944: gap of 100 bp
* 145945 178035: contig of 32091 bp in length.

Location/Qualifiers
1. 178035
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-22D13"
/clone_11b="RP11-11 Human Male BAC"

```
misc_feature 1..1212
              /note="assembly_fragment"
              clone_end:Sp6
              vector_side:right"
misc_feature 1313..5818
              /note="assembly_fragment"
              5919..9462
              /note="assembly_fragment"
              9563..15211
              /note="assembly_fragment"
              15312..28255
              /note="assembly_fragment"
              28356..44646
              /note="assembly_fragment"
              44747..63409
              /note="assembly_fragment"
              63510..90494
              /note="assembly_fragment"
              clone_end:T7
misc_feature 90595..118529
              /note="assembly_fragment"
              118630..145844
              /note="assembly_fragment"
              145945..178035
              /note="assembly_fragment"
BASE COUNT 48869 a 40236 c 39973 g 47951 t 1006 others
ORIGIN
```

```
Query Match 17.1%; Score 21; DB 2; Length 178035;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 98 ttatctcaattatccaca 118
|||||
Db 169635 TTATCTCAATTATTCACA 169655
```

```
RESULT 3
AC007600/c AC007600 183079 bp DNA HTG 03-JUL-2001
LOCUS Homo sapiens chromosome 16 clone RP11-264A16, WORKING DRAFT
DEFINITION
AC007600
AC007600.4 GI:14589421
VERSION HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 1 to 183079)
JOURNAL DOE Joint Genome Institute.
REFERENCE Sequencing of Human Chromosome 16
AUTHORS Unpublished
TITLE 2 (bases 1 to 183079)
JOURNAL Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E., Robinson, D.,
REFERENCE Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,
AUTHORS Bryant, J., Tesmer, J., Meinke, L., Longmire, J., White, S., Tatum, O.,
JOURNAL Campbell, C., Fawcett, J., Malbise, M., Bussod, M., Sutherland, R.,
REFERENCE McKury, K., Han, C. and Deaven, L.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-1999) Center for Human Genome Studies, DOE Joint
REFERENCE Genome Institute, Los Alamos National Laboratory, MS M888, Los
AUTHORS Alamos, NM 87545, USA
COMMENT On Jul 3, 2001 this sequence version replaced gi:8575868.
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
```

GenBank flat file format but are available as part of this entry's ASN.1 file.

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

* NOTE: This is a 'working draft' sequence. It currently consists of 1 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.

* This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

1 183079: contig of 183079 bp in length.

FEATURES

source 1..183079
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-264A16"

BASE COUNT 47478 a 41752 c 42357 g 51492 t

ORIGIN

```
Query Match 17.1%; Score 21; DB 2; Length 183079;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 98 ttatctcaattatccaca 118
|||||
Db 130958 TTATCTCAATTATTCACA 130938
```

```
RESULT 4
AC013298 AC013298 194198 bp DNA HTG 20-SEP-2000
LOCUS Homo sapiens clone RP11-3N3, WORKING DRAFT SEQUENCE, 19 unordered
DEFINITION
AC013298
AC013298.6 GI:10198404
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
TITLE 1 (bases 1 to 194198)
JOURNAL Birren, B., Linton, L., Nusbaum, C. and Lander, E.
REFERENCE 2 (bases 1 to 194198)
AUTHORS Unpublished
TITLE Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
REFERENCE Baldwin, J., Barna, N., Becker, R., Boguslavsky, L., Boukhalter, B.,
AUTHORS Brown, A., Castle, A., Colangelo, M., Collins, S., Collamore, A.,
REFERENCE Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
AUTHORS Ferreira, P., FitzHugh, W., Forrest, C., Funke, R., Gage, D.,
REFERENCE Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
AUTHORS Howland, J. C., Johnson, R., Jones, C., Kamp, L., Karatas, A., Klein, J.,
REFERENCE Lehoczy, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
AUTHORS McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
REFERENCE Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
AUTHORS Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
REFERENCE Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
AUTHORS Tesfaye, S., Tittell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
REFERENCE Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
```


*	1	16747:	contig of 16747 bp	in length
*	16748	16847:	gap of 100 bp	
*	16848	18184:	contig of 1337 bp	in length
*	18185	18284:	gap of 100 bp	
*	18285	21234:	contig of 2850 bp	in length
*	21235	21334:	gap of 100 bp	
*	21335	23492:	contig of 2158 bp	in length
*	23493	23592:	gap of 100 bp	
*	23593	26852:	contig of 3260 bp	in length
*	26853	26952:	gap of 100 bp	
*	26953	30877:	contig of 3925 bp	in length
*	30878	30977:	gap of 100 bp	
*	30978	35870:	contig of 4893 bp	in length
*	35871	35970:	gap of 100 bp	
*	35971	40651:	contig of 4721 bp	in length
*	40692	40791:	gap of 100 bp	
*	40792	44904:	contig of 4113 bp	in length
*	44903	45004:	gap of 100 bp	
*	45005	51760:	contig of 6756 bp	in length
*	51761	51860:	gap of 100 bp	
*	51861	58733:	contig of 6863 bp	in length
*	58724	58823:	gap of 100 bp	
*	58824	68160:	contig of 9337 bp	in length
*	68161	68260:	gap of 100 bp	
*	68261	84475:	contig of 16215 bp	in length
*	84476	84575:	gap of 100 bp	
*	84576	98534:	contig of 13959 bp	in length
*	98535	98534:	gap of 100 bp	
*	98635	114663:	contig of 16005 bp	in length
*	114664	114743:	gap of 100 bp	
*	114744	144676:	contig of 29933 bp	in length
*	144777	144776:	gap of 100 bp	
*	146048	160047:	contig of 15271 bp	in length
*	160148	160147:	gap of 100 bp	
*	160148	178836:	contig of 18689 bp	in length
*	178837	178936:	gap of 100 bp	
*	178937	194198:	contig of 15262 bp	in length
*	Location/Qualifiers			

```

source
1. .194198
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-3N3"
/clone_11b="RPc1-11 Human Male BAC"
1. .16747
/note="assembly-fragment
clone_end:SP6
vector_side:left"
16848..18184
/note="assembly-fragment"
18285..21254
/note="assembly-fragment"
21335..23492
/note="assembly-fragment"
23593..26852
/note="assembly-fragment"
26953..30877
/note="assembly-fragment"
30978..35870
/note="assembly-fragment"
35971..40691
/note="assembly-fragment"
40792..44904
/note="assembly-fragment"
45005..51760
/note="assembly-fragment"
51861..58773
/note="assembly-fragment"
58824..68160
/note="assembly-fragment"
68261..84475
/note="assembly-fragment"
84576..98534
/note="assembly-fragment"
98635..114643
/note="assembly-fragment"
114744..144676
/note="assembly-fragment"
144777..160047
/note="assembly-fragment"
160148..1718836
/note="assembly-fragment"
178937..194198

```

```

BASE COUNT      52345 a 44636 c 44383 g 51032 t 1802 others
ORIGIN
Query Match      17.1%; Score 21; DB 2; Length 194198;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      98 ttatcttcaatttattccaca 118
          |||||
Db 104368 TTATCTTCAATTATTATCCACA 104388

```

RESULT	5
AP000958/c	
LOCUS	
DEFINITION	AB000958 50974 bp DNA PRI 14-JAN-2000 Homo sapiens genomic DNA, chromosome 21q21.1-q21.2 clone:B794P4, IL56-AFP region, complete sequence.
ACCESSION	AF000958
VERSION	AP000958.2 GI:6705920
SOURCE	HGT.
KEYWORDS	Homo sapiens DNA, clone:B794P4. Homo sapiens
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	I (bases 1 to 50974)

AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Homo sapiens 50,974 genomic DNA of 21q21.1-q21.2
JOURNAL Published Only in Database (1999) In press
REFERENCE 2 (bases 1 to 50974)
AUTHORS Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P., Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1999) to the DDBJ/EMBL/GenBank databases.
REFERENCE Masahira Hattori, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
 (E-mail:hattori@gscc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923, Fax:81-42-778-9924)
 On Jan 16, 2000 this sequence version replaced gi:630658.
COMMENT Sequence updated (13-Jan-2000).
FEATURES Location/Qualifiers
source 1..50974
 /Organism="Homo sapiens"
 /db_xref="taxon:9606"
 /Chromosome="21"
 /Clone="B794P4"
 /map="21q21.1-q21.2"
BASE COUNT 16566 a 8657 c 8595 g 17156 t
ORIGIN
 Query Match 16.3%; Score 20; DB 9; Length 50974;
 Best Local Similarity 100.0%; Pred.No.1.3;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 97 ctatctcaattatcca 116
 ||||||||||||||||
 Db 21415 CTTACTTCATTATATCA 21396
RESULT 6
AC024713 155304 bp DNA 01-SEP-2000
LOCUS Homo sapiens chromosome 21 clone RP11-170F21, WORKING DRAFT
DEFINITION AC024713
ACCESSION AC024713.4 GI:9958179
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 155304)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 155304)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 On Sep 1, 2000 this sequence version replaced gi:8569835.
COMMENT ----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site:http://genome.wustl.edu/gsc/index.shtml
 ----- Project Information -----
 Center project name: H.NH0170F21
 ----- Summary Statistics -----
 Sequencing vector: M13; 100%
 Sequencing vector: plasmid; 0%
 Chemistry: Dye-Primer ET; 100% of reads
 Chemistry: Dye-terminator Big Dye; 0% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 138354 bases at least Q40

Consensus quality: 142506 bases at least Q30
 Consensus quality: 144038 bases at least Q20
 Insert size: 150000; agarose-fp
 Insert size: 150578; sum-of-ctctigs
 Quality coverage: 3.56 in Q20 bases; agarose-fp
 Quality coverage: 3.67 in Q20 bases; sum-of-ctctigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 31 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 1
 900 899: gap of 899 bp in length
 1000 999: gap of unknown length
 1826 1826: contig of 827 bp in length
 1827 1926: gap of unknown length
 1927 3373: contig of 1447 bp in length
 3374 3473: gap of unknown length
 3474 4709: contig of 1236 bp in length
 4710 4809: gap of unknown length
 4810 5988: contig of 1179 bp in length
 5989 6088: gap of unknown length
 6089 8166: contig of 2078 bp in length
 8167 8267 10820: contig of 2554 bp in length
 8267 10920: gap of unknown length
 10821 10920: gap of 1548 bp in length
 10921 12468: contig of 1548 bp in length
 12469 12568: gap of unknown length
 12569 14508: contig of 1940 bp in length
 14509 14608: gap of unknown length
 14609 16851: contig of 2243 bp in length
 16852 16951: gap of unknown length
 16952 19153: contig of 2202 bp in length
 19154 19253: gap of unknown length
 19254 22560: contig of 3307 bp in length
 22561 22660: gap of unknown length
 22661 25965: contig of 3305 bp in length
 25966 26065: gap of unknown length
 26066 30499: contig of 4434 bp in length
 30500 30599: gap of unknown length
 30600 33411: contig of 2812 bp in length
 33412 33511: gap of unknown length
 33512 37042: contig of 3531 bp in length
 37043 37142: gap of unknown length
 37143 40035: contig of 2893 bp in length
 40036 40135: gap of unknown length
 40136 45130: contig of 4995 bp in length
 45131 45230: gap of unknown length
 45231 49657: contig of 4427 bp in length
 49658 49757: gap of unknown length
 49758 54058: contig of 4301 bp in length
 54059 54158: gap of unknown length
 54159 58730: contig of 4572 bp in length
 58731 58830: gap of unknown length
 58831 64337: contig of 5507 bp in length
 64338 64437: gap of unknown length
 64439 69738: contig of 5301 bp in length
 69739 69838: gap of unknown length
 69839 76434: contig of 6596 bp in length
 76435 76534: gap of unknown length
 76535 84858: contig of 8324 bp in length
 84859 92762: gap of unknown length
 92763 92862: gap of unknown length
 92863 100409: contig of 7547 bp in length
 100410 100509: gap of unknown length
 100510 109405: contig of 8896 bp in length
 109406 109505: gap of unknown length
 121992 121992: contig of 12487 bp in length
 121993 122092: gap of unknown length

* 122093 134408: contig of 12316 bp in length
 * 134409 134508: gap of unknown length
 * 134509 155304: contig of 20796 bp in length.
 FEATURES
 source
 location/Qualifiers
 1.155304
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="21"
 /clone="RP11-170F21"
 1.899

misc_feature /note="assembly_name:Contig10"
 1000.1626
 /note="assembly_name:Contig9"
 1927.3373
 /note="assembly_name:Contig16"
 3474.4709
 /note="assembly_name:Contig17"
 4810.5988
 /note="assembly_name:Contig18"
 6089.8166
 /note="assembly_name:Contig19"
 8267.10820
 /note="assembly_name:Contig20"
 10921.12468
 /note="assembly_name:Contig21"
 12569.14508
 /note="assembly_name:Contig22"
 14609.16851
 /note="assembly_name:Contig23"
 16952.19153
 /note="assembly_name:Contig24"
 19254.22560
 /note="assembly_name:Contig25"
 22661.25965
 /note="assembly_name:Contig26"
 26066.30499
 /note="assembly_name:Contig27"
 30600.33411
 /note="assembly_name:Contig28"
 33512.37042
 /note="assembly_name:Contig29"
 37143.40035
 /note="assembly_name:Contig30"
 40136.45130
 /note="assembly_name:Contig31"
 45231.49657
 /note="assembly_name:Contig32"
 49758.54058
 /note="assembly_name:Contig33"
 54159.58730
 /note="assembly_name:Contig34"
 58831.64337
 /note="assembly_name:Contig35"
 64438.69738
 /note="assembly_name:Contig36"
 69839.76434
 /note="assembly_name:Contig37"
 76535.84858
 /note="assembly_name:Contig38"
 84959.92762
 /note="assembly_name:Contig39"
 92863.100409
 /note="assembly_name:Contig40"
 100510.109405
 /note="assembly_name:Contig41"
 109506.121992
 /note="assembly_name:Contig42"
 122093.134408
 /note="assembly_name:Contig43"
 134509.155304
 /note="assembly_name:Contig44"
 misc_feature 50522 a 26013 c 26393 g 49371 t 3005 others
 BASE COUNT
 ORIGIN

Query Match 16.3%; Score 20; DB 2; Length 155304;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 97 ctatctcaattatcca 116
 ||||||||||||||||
 Db 47722 CTATCTTCATTATTC 47741

RESULT 7
 CDS07EPA/c
 LOCUS
 DEFINITION
 PROGRESS *** 2 ordered pieces.
 ACCESSION
 AL512361
 VERSION
 AL512361.2 GI:13548503
 KEYWORDS
 HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT.
 SOURCE
 human.
 ORGANISM
 Homo sapiens

REFERENCE
 AUTHORS
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 Heilig, R., Petit, J.L., Vico, V., Dasilva, C., Robert, C., Mincker, P.,
 Brothier, P., Catolico, L., Barbe, V., Pelletier, E., Artiguenave, F.,
 Levy, M., Eckenberg, R., Bruls, T., deBerardinis, V., Cruaud, C.,
 Gyapay, G., Saurin, W. and Weissensbach, J.
 Sequencing of the human chromosome 14
 Unpublished
 2 (bases 1 to 158927)
 Genoscope.
 Direct Submission
 Submitted (29-MAY-2001) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)
 - Web : www.genoscope.cns.fr)
 On Apr 4, 2001 this sequence version replaced gi:11990876.

COMMENT

Overall quality chart :
 Range : bases
 0 : 598
 1 - 9 : 17
 10 - 19 : 203
 20 - 29 : 567
 30 - 39 : 1808
 40 - 49 : 5526
 50 - 59 : 7642
 60 - 69 : 15666
 70 - 79 : 36142
 80 - 89 : 53454
 90 - 99 : 37304

Percentage of bases with a quality value >= 40 : 97 %

 Genome Center
 Center: Genoscope / Centre National de Sequencage
 Center code: GS
 Web site: http://www.genoscope.cns.fr/
 Contact: sequef@genoscope.cns.fr

The following BAC sequence is oriented from the T7 to the SP6 end.
 IMPORTANT: This sequence is unfinished and does not necessarily
 represent the correct sequence. Work on the sequence is in progress
 and the release of this data is based on the understanding that the
 sequence may change as work continue. The sequence may be
 contaminated with foreign sequence from E.coli, yeast, vector,
 phage, etc. . even if efforts are made to eliminate these
 contaminating sequences.
 Upstream BAC (overlapping the T7 end) : C-3172B5
 Downstream BAC (overlapping the SP6 end) : R-346L24 (AC-AL163612)

 Summary Statistics
 Assembly program: Phrap, version 2.0
 Quality coverage: 5.91x in Q20 bases; sum-of-contigs

Contigs composition :
 98062 bp contig from 1 to 98062
 60765 bp contig from 98163 to 158927.
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submittor.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 98062: contig of 98062 bp in length
 * 98063 98162: gap of 100 bp
 * 98163 158927: contig of 60765 bp in length.
 Location/Qualifiers
 1..158927
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="14"
 /clone_lib="RPC1-11"
 /clone_1lb="RPC1-11"
 77210..77462
 /note="matching EMBL:Z66661
 RHdb:RH53692
 RHdb:RH53795
 dbSTS:STS8872
 Identified using the e-PCR software (G. Schuler)"
 89489..89715
 /note="matching EMBL:Z23897
 RHdb:RH53481
 RHdb:RH53636
 dbSTS:STS9781
 Identified using the e-PCR software (G. Schuler)"
 89541..89698
 /note="matching EMBL:Z23897
 RHdb:RH95620
 RHdb:RH49080
 RHdb:RH13032
 RHdb:RH1093
 dbSTS:STS1260
 Identified using the e-PCR software (G. Schuler)"
 BASE COUNT 49681 a 32396 c 30501 g 46179 t 170 others
 ORIGIN

Query Match 16.3% Score 20; DB 2; Length 158927;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 93 gatactatctcaattat 112
 |||||
 Db 148122 GATACTATCTCAATTAT 148103

RESULT 8
 CNS01R11/c 171345 bp DNA PRI 28-APR-2001
 LOCUS Human chromosome 14 DNA sequence BAC R-346L24 of library RPC1-11
 DEFINITION from chromosome 14 of Homo sapiens (human), complete sequence.
 ACCESSION AL163612.5 GI:13276126
 VERSION
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 171345)
 Hellig,R., Pettit,J.L., Vico,V., Dasilva,C., Robert,C., Wincker,P.,
 Brothier,P., Catolico,L., Barbe,V., Pelleter,E., Artiguenave,F.,
 Levy,M., Eckenberg,R., Bruls,T., deBerradins,V., Cruaud,C.,
 Gysay,G., Saurin,W. and Weissenbach,J.
 Sequencing of the human chromosome 14

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 171345)
 AUTHORS Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (26-APR-2001) Genoscope - Centre National de Sequencage :
 BP 191 91006 Evry cedex - FRANCE (E-mail : sequef@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 On Mar 12, 2001 this sequence version replaced g1:11875958.
 Center: Genoscope / Centre National de Sequencage
 Center code: GS
 Web site: http://www.genoscope.cns.fr/
 Contact: sequef@genoscope.cns.fr

COMMENTS
 The following BAC sequence is oriented from the T7 to the SP6 end.
 Upstream BAC (overlapping the T7 end) : R-326E7
 Downstream BAC (overlapping the SP6 end) : R-649E7 (AC-AL139099)
 ----- Summary Statistics -----
 Assembly program: Phrap; version 2.0
 Quality coverage: 9.09x in Q20 bases; sum-of-contigs

 Overall quality chart :
 Range : bases
 0 :
 1 - 9 :
 10 - 19 : 2
 20 - 29 : 9
 30 - 39 : 186
 40 - 49 : 2239
 50 - 59 : 7900
 60 - 69 : 8553
 70 - 79 : 12847
 80 - 89 : 31205
 90 - 99 : 108404

Percentage of bases with a quality value >= 40 : 99 %.

FEATURES
 source
 1..171345
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="14"
 /clone_lib="RPC1-11"
 /clone_1lb="RPC1-11"

BASE COUNT 55535 a 35560 c 32962 g 47288 t
 ORIGIN

Query Match 16.3% Score 20; DB 9; Length 171345;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 93 gatactatctcaattat 112
 |||||
 Db 25550 GATACTATCTCAATTAT 25531

RESULT 9
 AC024417 182340 bp DNA HTG 12-MAR-2000
 LOCUS Homo sapiens chromosome 4 clone RP11-221C21 map 4, *** SEQUENCING
 DEFINITION IN PROGRESS ***; 64 unordered pieces.
 ACCESSION AC024417
 VERSION AC024417.2 GI:7229920
 KEYWORDS
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 182340)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens chromosome 4, clone RP11-221C21
 Unpublished

REFERENCE
AUTHORS

2 (bases 1 to 182340)
Birtren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Boukhalter,B., Brown,A., Burkett,G., Campolano,A., Castle,A.,
Choepep,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Deatellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
Galaan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Kartas,A.,
Klein,J., Landers,T., Larcocque,K., Lehoczeky,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrum,J.,
Meneus,L., Mihova,T., Miranda,C., Mienga,Y., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,
Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A.,
Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.

TITLE
JOURNAL
COMMENT

Direct Submission
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 12, 2000 this sequence version replaced gi:7108213.
All repeats were identified using RepeatMasker:

Smith, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L6906

Center clone name: 221_C_21

* NOTE: This is a 'working draft' sequence. It currently
* consists of 64 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1056 1155: contig of 1055 bp in length
1156 2524: contig of 1369 bp in length
2525 2624: gap of 100 bp
2625 3768: contig of 1144 bp in length
3769 3868: gap of 100 bp
3869 4906: contig of 1038 bp in length
4907 5006: gap of 100 bp
5007 6012: contig of 1006 bp in length
6013 6112: gap of 100 bp
6113 7527: contig of 1415 bp in length
7528 7627: gap of 100 bp
7628 8685: contig of 1238 bp in length
8686 8965: gap of 100 bp
8966 9989: contig of 1024 bp in length
9990 10089: gap of 100 bp
10090 11332: contig of 1243 bp in length
11333 11432: gap of 100 bp
11433 12597: contig of 1165 bp in length
12598 12697: gap of 100 bp
12698 13979: contig of 1282 bp in length
13980 14079: gap of 100 bp
14080 15544: contig of 1465 bp in length
15545 15644: gap of 100 bp
15645 17090: contig of 1446 bp in length
17091 17190: gap of 100 bp
17191 18508: contig of 1318 bp in length

18509 18608: gap of 100 bp
18609 19829: contig of 1221 bp in length
19830 19929: gap of 100 bp
19930 21213: contig of 1284 bp in length
21214 21313: gap of 100 bp
21314 22808: contig of 1495 bp in length
22809 22908: gap of 100 bp
22909 24534: contig of 1626 bp in length
24535 24634: gap of 100 bp
24635 27162: contig of 2528 bp in length
27163 27262: gap of 100 bp
27263 28811: contig of 1549 bp in length
28812 28911: gap of 100 bp
28912 30046: contig of 1135 bp in length
30047 30146: gap of 100 bp
30147 31304: contig of 1158 bp in length
31305 31404: gap of 100 bp
31405 33474: contig of 2070 bp in length
33475 33574: gap of 100 bp
33575 35085: contig of 1511 bp in length
35086 35185: gap of 100 bp
35186 36846: contig of 1661 bp in length
36847 36946: gap of 100 bp
36947 38786: contig of 1840 bp in length
38787 38886: gap of 100 bp
38887 41015: contig of 2129 bp in length
41016 41115: gap of 100 bp
41116 43401: contig of 2286 bp in length
43402 43501: gap of 100 bp
43502 44790: contig of 1289 bp in length
44791 44890: gap of 100 bp
44891 46931: contig of 2041 bp in length
46932 47031: gap of 100 bp
47032 50023: contig of 2992 bp in length
50024 50123: gap of 100 bp
50124 52810: contig of 2687 bp in length
52811 52910: gap of 100 bp
52911 55013: contig of 2103 bp in length
55014 55113: gap of 100 bp
55114 57694: contig of 2581 bp in length
57695 57794: gap of 100 bp
57795 60148: contig of 2354 bp in length
60149 60248: gap of 100 bp
60249 62576: contig of 2328 bp in length
62577 62676: gap of 100 bp
62677 64748: contig of 2072 bp in length
64749 64848: gap of 100 bp
64849 67425: contig of 2577 bp in length
67426 67525: gap of 100 bp
67526 70617: contig of 3092 bp in length
70618 70717: gap of 100 bp
70718 72665: contig of 1948 bp in length
72666 72765: gap of 100 bp
72766 75102: contig of 2337 bp in length
75103 75202: gap of 100 bp
75203 78445: contig of 3243 bp in length
78446 78545: gap of 100 bp
78546 81397: contig of 2852 bp in length
81398 81497: gap of 100 bp
81498 84343: contig of 2846 bp in length
84344 84443: gap of 100 bp
84444 86480: contig of 2037 bp in length
86481 86580: gap of 100 bp
86581 88846: contig of 2266 bp in length
88847 88946: gap of 100 bp
88947 93512: contig of 4566 bp in length
93513 93612: gap of 100 bp
93613 96846: contig of 3234 bp in length
96847 96946: gap of 100 bp
96947 99557: contig of 2611 bp in length
99558 99657: gap of 100 bp
99658 102370: contig of 2713 bp in length
102371 102470: gap of 100 bp

```

* 102471 106107: contig of 3637 bp in length
* 106108 106207: gap of 100 bp
* 106208 109111: contig of 2904 bp in length
* 109112 109211: gap of 100 bp
* 109212 113450: contig of 4239 bp in length
* 113451 113550: gap of 100 bp
* 113551 118394: contig of 4844 bp in length
* 118395 118494: gap of 100 bp
* 118495 123220: contig of 4726 bp in length
* 123221 123320: gap of 100 bp
* 123321 127732: contig of 4412 bp in length
* 127733 127832: gap of 100 bp
* 127833 132756: contig of 4924 bp in length
* 132757 132856: gap of 100 bp
* 132857 137254: contig of 4398 bp in length
* 137255 137354: gap of 100 bp
* 137355 143801: contig of 6447 bp in length
* 143802 143901: gap of 100 bp
* 143902 148857: contig of 4956 bp in length
* 148858 148957: gap of 100 bp
* 148958 156185: contig of 7228 bp in length
* 156186 156285: gap of 100 bp
* 156286 164408: contig of 8123 bp in length
* 164409 164508: gap of 100 bp
* 164509 171697: contig of 7189 bp in length
* 171698 171797: gap of 100 bp
* 171798 182340: contig of 10543 bp in length.

```

FEATURES
source
1. 182340
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-221C21"
/clone_lib="RPC1-11 Human Male BAC"
1. 1055
/note="assembly_fragment"
1156. 2524
/note="assembly_fragment"
misc_feature
2625. 3768
misc_feature

Query Match 16.3%: Score 20; DB 2; Length 182340;
Best Local Similarity 100.0%; Pred.No. 1.1;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 97 ctatctcaattattcca 116
|||||
Db 178749 CTTATCTCAATTATTCCA 178768

RESULT 10
AP001675/c DNA PRI 30-MAY-2000
LOCUS Homo sapiens genomic DNA, chromosome 21q, section 19/105.
DEFINITION AP001675 AL163220 BA000005
ACCESSION AP001675.1 GI:7768691
VERSION
KEYWORDS
SOURCE Homo sapiens DNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (sites)
REFERENCE Hattori,M., Fujiiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,
AUTHORS Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M. and Schudy,A.
TITLE The DNA sequence of human chromosome 21. The chromosome 21 mapping
and sequencing consortium
JOURNAL Nature 405. (6784), 311-319 (2000)
MEDLINE 20289799
REFERENCE 2 (bases 1 to 340000)
AUTHORS Hattori,M., Fujiiyama,A., Taylor,T.D., Watanabe,H., Yada,T.,

COMMENT

TITLE
JOURNAL

Park,H.S., Toyoda,A., Ishii,K., Totoki,Y., Choi,D.K., Soeda,E.,
Ohki,M., Takagi,T., Sakaki,Y., Taudien,S., Blechschmidt,K.,
Polley,A., Menzel,U., Delabar,J., Kumpf,K., Lehmann,R.,
Patterson,D., Reichwald,K., Rump,A., Schillhabel,M., Schudy,A.,
Zimmermann,W., Rosenthal,A., Kudoh,J., Shibuya,K., Kawasaki,K.,
Asakawa,S., Shintani,A., Sasaki,T., Nagamine,K., Mitsuyma,S.,
Antonarakis,S.E., Minoshima,S., Shimizu,N., Nordstrek,G.,
Hornischer,K., Barand,P., Scharfe,M., Schoen,O., Desario,A.,
Reichelt,J., Kauer,G., Bloeker,H., Ramser,J., Beck,A., Klages,S.,
Hennig,S., Risseilmann,L., Dagnad,E., Wehrmeyer,S., Borzym,K.,
Gardner,K., Nizetic,D., Francis,F., Lehnach,H., Reinhardt,R. and
Vaspo,M.L.

Direct Submission
Submitted (10-Apr-2000) to the DDBJ/EMBL/GenBank databases. The
Chromosome 21 Mapping and Sequencing Consortium: * RIKEN Genomic
Sciences Center, Human Genome Research Group * Institute of
Molecular Biotechnology, Genome Analysis * Keio University School
of Medicine, Dept. of Molecular Biology * GBF, Dept. of Genome
Analysis * Max-Planck Institute for Molecular Genetics (addresses
see below)

On May 30, 2000 this sequence version replaced gi:7717272.
The chromosome 21 mapping and sequencing consortium consisting of
* RIKEN Genomic Sciences Center, Human Genome Research Group, *
Sagamihara 228-8555, Japan,
* e.mail: hattori@gscl.riken.go.jp/
* URL: http://hgp.gscl.riken.go.jp/
and
* Institute of Molecular Biotechnology, Genome Analysis, *
Beutenbergstrasse 11, D-07745 Jena, Germany,
* e.mail: gscj-submit@genome.imb-jena.de
* URL: http://genome.imb-jena.de/
and
* Keio University School of Medicine, Molecular Biology, * Tokyo
160-8582, Japan,
* e.mail: nshimiz@dm.med.keio.ac.jp/
* URL: http://www.dmb.med.keio.ac.jp/
and
* GBF, Dept. of Genome Analysis,
* Mascheroder Weg 1, D-38124 Braunschweig, Germany, * e.mail:
info.genome@gbf.de
* URL: http://genome.gbf.de/
and
* Max-Planck Institute for Molecular Genetics,
* Innesstrasse 73, D-14195 Berlin, Germany,
* e.mail: info-chr21@molgen.mpg.de
* URL: http://chr21.rz-berlin.mpg.de/
AL163220: Submitted (10-Apr-2000).

FEATURES

source
1. 340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
<1. 9371
/note="Accession No. AL132993"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="C103E2331, 5' partial"
/clone_lib="TCRFc103 Cosmid library"
<1. 113120
/note="Accession No. AP000431"
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q21.1"
/clone="R28W9, 5' partial"
/clone_lib="RPC1-11 BAC library"
complement(1510..1686)
/note="LIMD1"
/rpt_family="LINE/L1"
/rpt_type="DISPERSED"
repeat_region
complement(1719..1947)
repeat_region

```

/note="L1M4"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
repeat_region complement(1948..2340)
/note="L1MD"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
repeat_region 4085..4116
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
repeat_region 4117..4158
/note="(TTTGG)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region 4159..4309
/note="MIR"
/rpt_family="SINE/MIR"
/rpt_type=DISPERSED
repeat_region 4336..5118
/note="L1P5"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
repeat_region complement(5588..6010)
/note="MLT1H"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 6279..6642
/note="MSTA"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 6648..7011
/note="MSTA-internal"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region complement(7020..11228)
/note="L1PA2"
/rpt_family="LINE/L1"
/rpt_type=DISPERSED
repeat_region 11194..11962
/note="MSTA-internal"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 11963..12333
/note="THE1B"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 13901..14287
/note="THE1A"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 14288..14493
/note="MSTA-internal"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 14494..15181
/note="LTR8"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
repeat_region 15182..15519
/note="MSTA-internal"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 15522..15620
/note="MSTA"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region complement(15621..16036)
/note="MSTA"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 16091..16223
/note="L1MA1"

```

```

/rpt_family="LINE/L1"
/rpt_type=DISPERSED
repeat_region 16256..16529
/note="MSTA"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 17192..17214
/note="(TTTAA)n"
/rpt_family="Simple_repeat"
/rpt_type=TANDEM
repeat_region complement(17215..17497)
/note="Alusg"
/rpt_family="SINE/Alu"
/rpt_type=DISPERSED
repeat_region complement(18031..18355)
/note="THE1B"
/rpt_family="LTR/MaLR"
/rpt_type=DISPERSED
repeat_region 18356..18783
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
repeat_region 18946..19195
/note="L2"
/rpt_family="LINE/L2"
/rpt_type=DISPERSED
repeat_region 19247..19419
/note="LTR35A"
/rpt_family="LTR/Retroviral"
/rpt_type=DISPERSED
repeat_region 19462..19540
/note="Alusg/x"
/rpt_family="SINE/Alu"

```

Query Match 16.3%; Score 20; DB 9; Length 340000;
 Best Local Similarity 100.0%; Pred. No. 1;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 ctatcttcaattatcca 116
 |||||
 Db 329060 CTTATCTCAATTATTCa 329041

```

RESULT 11
LOCUS U41746 27710 bp DNA INV 10-JUL-2001
DEFINITION Caenorhabditis elegans cosmid T18H9, complete sequence.
ACCESSION U41746
VERSION U41746.1 GI:1118132
KEYWORDS HTG..
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 27710)
AUTHORS Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
TITLE Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 27710)
AUTHORS The C. elegans Sequencing Consortium.
TITLE Genome sequence of the nematode C. elegans: a platform for
JOURNAL investigating biology. The C. elegans Sequencing Consortium
MEDLINE Science 282 (5396), 2012-2018 (1998)
REFERENCE 2 (bases 1 to 27710)
AUTHORS Gattung S. and Du Z.
TITLE The sequence of C. elegans cosmid T18H9
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 27710)
REFERENCE 3 (bases 1 to 27710)
AUTHORS Waterston R.
TITLE Direct Submission
JOURNAL Unpublished
AUTHORS 4 (bases 1 to 27710)
REFERENCE 4 (bases 1 to 27710)
AUTHORS Waterston R.
TITLE Direct Submission
JOURNAL Submitted (30-NOV-1995) Robert Waterston
REFERENCE 5 (bases 1 to 27710)

```


VSLSQELCSILGSSMSPEGSDAVWLCNLSMPAYSIIGGPVPIVNCPSFOVROTLSH
VIRIONFCHMSNSNPETIVGILIGTDFSOVKAIVYCLAKHSKIAMPVNHLPFAI
PSTSLFYKMGICIDPOLNCRDLMDREGDMTLPEKTSIANKIAMPVNSSKKAM
PIGEAMQLTADBDPEGGRVDFPLESEVRGSOQCEIENDDDNGNGRAQVITSPRE
ESVISTKDGSSPHSPQLRNVQELNMDPVPNPVPOSSOYSGDNGPSSNTNPN
SKKDTAKVTKIAHFKTLVSRSPSGSLSTLYLKEKRDMLKMLCKMKGOKPREG
AAPSQVSSIGRMICATGKHNELTVIIDGNAYNGIRYFQTSQWQHIKSFIAFTPT
MVCTA"

gene
25268.. 27631
/gene="T18H9.1"
/note="gid-6"
join(25268..25334,25502..25749,25795..26001,26180..26566,
26617..26915,26956..27037,27086..27411,27487..27631)
/gene="T18H9.1"
/note="coded for by the following C. elegans CDNAs:
yk68a8.5"
/codon_start=1
/product="Hypothetical protein T18H9.1"
/protein_id="AAA8334.1"
/db_xref="GI:118137"
/translation="MPSIRYRLVALVVFISVYGOEAAVQTPPYQPNILREGSG
EPYRAIRPNQOLQPDIDPLRDERKNNNAGYIRHPTFVPOQOQFRPAGVY
IAQPNRVYSQTFKSAVVRVPIVFOQFOKQGSQYIQQITRQOQOYSQPRFL
RQQVAPRVAVYIAGRLTGPIAPVYQGSERPAVAVIERPVAPRTPIERPV
ARPAPIERPEAPAPVIERPVAPVPIEPTPARPAPVIEPTAKQPPQPPRT
RPVAPSTTORIETPTQPTTRATKRTTTTAAPTPRLTARATPLATTSR

Query Match 15.4%; Score 19; DB 3; Length 27710;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 98 ttatctcaattattcca 116
|||||
Db 23408 TTATCTTCATTTATTCCA 23390

RESULT 12
AL357912
LOCUS
DEFINITION Human DNA sequence from clone RP11-328N1 on chromosome 1, complete
sequence.
ACCESSION AL357912 AC022524
VERSION AL357912.10 GI:14018273
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 130057)
AUTHORS Hall, R.
TITLE Direct Submission
JOURNAL Submitted (09-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On May 15, 2001 this sequence version replaced gi:8569932
gi:13751415.

COMMENT
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em; EMBL; Sw;
SWISSPROT; Tr; TrEMBL; Wp; WormPep; Information on the WormPep
database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP11-328N1 is from the library RPCT-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-328N1. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP11-328N1 is at 1 in this sequence. The
true left end of clone RP5-1102M4 is at 129958 in this sequence.

FEATURES
source

1..130057
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-328N1"
/clone_1lb="RPCT-11.2"
3..166
/note="MSTC repeat: matches 1..168 of consensus"
171..1394
/note="LM1C repeat: matches 321..1600 of consensus"
1431..1870
/note="LIM4 repeat: matches 3258..3689 of consensus"
1871..2174
/note="AluSx repeat: matches 1..306 of consensus"
2175..3813
/note="LIM4 repeat: matches 3689..5371 of consensus"
3814..8354
/note="LIM4 repeat: matches 1499..6145 of consensus"
8365..8676
/note="LIM4 repeat: matches 5814..6123 of consensus"
8685..8990
/note="AluSx repeat: matches 1..291 of consensus"
9080..9234
/note="HAI1 repeat: matches 1523..1692 of consensus"
10201..10493
/note="AluSx repeat: matches 1..293 of consensus"
10626..10923
/note="AluJb repeat: matches 1..297 of consensus"
12090..12320
/note="MER20 repeat: matches 1..218 of consensus"
13930..14603
/note="LIM4 repeat: matches 5620..6304 of consensus"
14617..20083
/note="LIP4 repeat: matches 649..6139 of consensus"
20157..20374
/note="LRR9 repeat: matches 427..644 of consensus"
20376..20630
/note="LIM4 repeat: matches 5376..5630 of consensus"
20631..21180
/note="LRR28 repeat: matches 1..448 of consensus"
21183..27754
/note="LIM4 repeat: matches 1387..5378 of consensus"
27850..27996
/note="LIM4 repeat: matches 238..386 of consensus"
28485..29096
/note="L2 repeat: matches 2064..2709 of consensus"
29211..29319
/note="L2 repeat: matches 1687..1795 of consensus"
29487..29592
/note="L2 repeat: matches 1335..1441 of consensus"
30157..30218
/note="31 copies 2 mer tt 67% conserved"
31310..31624
/note="AluSx repeat: matches 1..311 of consensus"
32682..33408
/note="LRR8 repeat: matches 1..691 of consensus"
33523..33812
/note="LIM4 repeat: matches 3952..4265 of consensus"

```

repeat_region 33829..34304
/note="L1MA7 repeat: matches 5815..6289 of consensus"
repeat_region 34305..34552
/note="AlusX repeat: matches 38..285 of consensus"
repeat_region 34556..34605
/note="25 copies 2 mer aa 72% conserved"
repeat_region 36354..36494
/note="TIGGR1 repeat: matches 2278..2418 of consensus"
repeat_region 36495..36761
/note="AlusX repeat: matches 1..267 of consensus"
repeat_region 36762..38999
/note="TIGGR1 repeat: matches 49..2278 of consensus"
repeat_region 39000..40970
/note="L1PA3 repeat: matches 1596..3569 of consensus"
repeat_region 40964..43533
/note="L1PA3 repeat: matches 3578..6146 of consensus"
repeat_region 43561..43620
/note="TIGGR1 repeat: matches 1..60 of consensus"
repeat_region 43899..44415
/note="MTD repeat: matches 1..505 of consensus"
repeat_region 44915..50056
/note="L1PA1 repeat: matches 951..6163 of consensus"
repeat_region 50057..50356
/note="AlusX repeat: matches 9..308 of consensus"
repeat_region 50357..50428
/note="L1PA1 repeat: matches 880..951 of consensus"
repeat_region 50410..51676
/note="L1PA1 repeat: matches 1422..750 of consensus"
repeat_region 52792..53029
/note="L2 repeat: matches 1740..1982 of consensus"
repeat_region 53427..53456
/note="15 copies 2 mer ta 86% conserved"
repeat_region 53883..54161
/note="MER44A repeat: matches 57..330 of consensus"
repeat_region 55356..55645
/note="AlusX repeat: matches 1..293 of consensus"
repeat_region 57603..57720
/note="59 copies 2 mer ag 71% conserved"
repeat_region 59372..59459
/note="MER91B repeat: matches 90..180 of consensus"
repeat_region 60347..61105
/note="L1PA3 repeat: matches 5397..6156 of consensus"
repeat_region 61106..61408
/note="AlusX repeat: matches 1..303 of consensus"
repeat_region 61409..63432
/note="L1PA3 repeat: matches 3326..5397 of consensus"
repeat_region 63446..63718
/note="AlusX repeat: matches 3..292 of consensus"
repeat_region 65907..66088
/note="MER5A repeat: matches 1..183 of consensus"
repeat_region 67083..67456
/note="L1 repeat: matches 4640..5052 of consensus"
repeat_region 67669..67874
/note="L1PA4 repeat: matches 5319..5531 of consensus"
repeat_region 68960..69182
/note="MTD repeat: matches 234..466 of consensus"
repeat_region 69306..69415
/note="MTD repeat: matches 124..226 of consensus"
repeat_region 69416..69723
/note="AlusX repeat: matches 1..307 of consensus"
repeat_region 69724..69811
/note="MTD repeat: matches 16..124 of consensus"
repeat_region 70639..70787
/note="L1MA6 repeat: matches 6027..6183 of consensus"
repeat_region 73464..73779
/note="L1MA4 repeat: matches 2938..3264 of consensus"
repeat_region 73790..75563
/note="L1MA2 repeat: matches 4311..6108 of consensus"
repeat_region 75564..75876
/note="AlusX repeat: matches 1..312 of consensus"
repeat_region 75877..76621
/note="L1MA2 repeat: matches 3554..4311 of consensus"
repeat_region 78308..78593

```

```

repeat_region /note="AlusB repeat: matches 1..287 of consensus"
80898..86972
/note="L1PA4 repeat: matches 11..6144 of consensus"
repeat_region 86996..88410
/note="L1 repeat: matches 3108..4548 of consensus"
repeat_region 88411..88491
/note="Alus repeat: matches 222..302 of consensus"
repeat_region 88499..88549
/note="TIGR1 repeat: matches 136..195 of consensus"
repeat_region 88550..88762
/note="AlusX repeat: matches 1..216 of consensus"
repeat_region 88763..89267
/note="L1 repeat: matches 2527..3102 of consensus"
repeat_region 89269..89605
/note="TIGR1 repeat: matches 1..364 of consensus"
repeat_region 89606..90914
/note="TIGR1-INTERNAL repeat: matches 1..1354 of consensus"
repeat_region 90916..91268

```

```

Query Match 15.4%; Score 19; DB 9; Length 130057;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 97 ctatctcaattatctcc 115
Db 77682 CTATCTCAATTATTC 77700

```

RESULT 13

```

AC027498/c AC027498 180911 bp DNA HTG 20-APR-2000
LOCUS Homo sapiens chromosome 18 clone RP11-656L11 map 18, WORKING DRAFT
DEFINITION
SEQUENCE, 25 unordered pieces.
AC027498
AC027498.2 GI:7622437
VERSION
KEYWORDS
HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE
human.
ORGANISM
Homo sapiens

```

```

REFERENCE
1 (bases 1 to 180911)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Balgutin,J., Barna,N., Bastien,V., Bedalov,F.,
Boguslavsky,L., Boucknight,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choeilano,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Dearlano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glende,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,U., Lacombe,K., Lamazares,R., Landers,T., Lehotzky,U.,
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Margulis,N.,
McCarthy,M., McEwan,P., McGuire,A., McKernan,K., McPherson,R.,
Meldrum,J., Menus,L., Mihova,T., Miranda,C., Mlewa,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testfay,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 20, 2000 this sequence version replaced gi:7342242.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

```

```

TITLE
JOURNAL
COMMENT

```

```

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8451
Center clone name: 656.L.11
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 165132 bases at least Q40
Consensus quality: 173168 bases at least Q30
Consensus quality: 176502 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 178511; sum-of-ctrls
Quality coverage: 3.9 in Q20 bases; agarose-fp
Quality coverage: 4.0 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 25 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
2315: contig of 2315 bp in length
2316 2415: gap of 100 bp
2416 6148: contig of 3733 bp in length
6149 6248: gap of 100 bp
6249 9065: contig of 2817 bp in length
9066 9165: gap of 100 bp
9166 13447: contig of 4282 bp in length
13448 13547: gap of 100 bp
13548 17515: contig of 3968 bp in length
17516 17615: gap of 100 bp
17616 21811: contig of 4196 bp in length
21812 21911: gap of 100 bp
21912 26026: contig of 4115 bp in length
26027 26126: gap of 100 bp
26127 29586: contig of 3460 bp in length
29587 29686: gap of 100 bp
29687 34675: contig of 4989 bp in length
34676 34775: gap of 100 bp
34776 39152: contig of 4377 bp in length
39153 39252: gap of 100 bp
39253 44142: contig of 4890 bp in length
44143 44242: gap of 100 bp
44243 48561: contig of 4319 bp in length
48562 48661: gap of 100 bp
48662 53987: contig of 5326 bp in length
53988 54087: gap of 100 bp
54088 60371: contig of 6284 bp in length
60372 60471: gap of 100 bp
60472 67910: contig of 7439 bp in length
67911 68010: gap of 100 bp
68011 74138: contig of 6128 bp in length
74139 74238: gap of 100 bp
74239 80142: contig of 5904 bp in length
80143 80242: gap of 100 bp
80243 88060: contig of 7818 bp in length
88061 88160: gap of 100 bp
88161 95098: contig of 6938 bp in length
95099 95198: gap of 100 bp
95199 104637: contig of 9439 bp in length
104638 104737: gap of 100 bp
104738 114963: contig of 10246 bp in length
114964 115083: gap of 100 bp
115084 124831: contig of 9748 bp in length
124832 124931: gap of 100 bp

```

```

FEATURES
Source
124932 136857: contig of 11926 bp in length
136858 136957: gap of 100 bp
136958 150860: contig of 13903 bp in length
150861 150960: gap of 100 bp
150961 180911: contig of 29951 bp in length.
Location/Qualifiers
1. 180911
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="18"
/map="18"
/clone="RP11-656L11"
/clone_11p="RPC1-11 Human Male BAC"
1. 2315
/note="assembly-fragment"
2416. 6148
/note="assembly-fragment"
6249. 9065
/note="assembly-fragment"
9166. 13447
/note="assembly-fragment"
13548. 17515
/note="assembly-fragment"
17616. 21811
/note="assembly-fragment"
21912. 26026
/note="assembly-fragment"
26127. 29586
/note="assembly-fragment"
29687. 34675
/note="assembly-fragment"
34776. 39152
/note="assembly-fragment"
39253. 44142
/note="assembly-fragment"
44243. 48561
/note="assembly-fragment"
48662. 53987
/note="assembly-fragment"
54088. 60371
/note="assembly-fragment"
60472. 67910
/note="assembly-fragment"
68011. 74138
/note="assembly-fragment"
74239. 80142
/note="assembly-fragment"
80243. 88060
/note="assembly-fragment"
88161. 95098
/note="assembly-fragment"
95199. 104637
vector_side:left"
104738. 114983
/note="assembly-fragment"
115084. 124831
/note="assembly-fragment"
124932. 136857
/note="assembly-fragment"
clone_end:Spe
vector_side:right"
136958. 150860
/note="assembly-fragment"
150961. 180911
/note="assembly-fragment"
BASE COUNT 58145 a 33104 c 32922 g 54339 t 2401 others
ORIGIN
Query Match 15.4%; score 19; DB 2; length 180911;
Best Local Similarity 100.0%; Pred. No. 4.1;

```


gene
 CDS
 complement(4819. .5154)
 /gene="NMB0393"
 /complement(4819. .5154)
 /note="similar to GP:3647360 percent identity: 78.30; identified by sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="multidrug resistance protein"
 /protein_id="AA040833.1"
 /db_xref="GI:7225615"
 /translation="MQMHWLFLTVAILSEVCGSSMLKSGSEKMPISGVVFSVC
 FMAISMTLKTPLATAVAIWAGVGLVITALVVFGEKADFIGIVSIGILLGVLL
 NIMSHSGH"
 complement(5306. .6418)
 /gene="NMB0394"
 /complement(5306. .6418)
 /note="similar to SP:P11458 GB:X12713 PID:581137 GB:000096
 PID:1651334 percent identity: 82.02; identified by
 sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="guinolate synthetase A"
 /protein_id="AA040834.1"
 /db_xref="GI:7225616"
 /translation="MOTAAARSFDYDMLIOTPTSACQIQAMAKVADPPRETADRL
 KDEIKALKKKNVAVAHVYVDFDIODLAEKGSGDSEMARFGHEHAGGLVAG
 VRKSGSAKTCPEKTYLMDLAECSLDGCEBEASACDHPRTYVYVANTSA
 VKRADVVTSSVALEIVSTLKSREKLTGEPDRHLDYICRETDMLMOSCTIV
 NEFGQELALAKHEPAVLVHPESPQSYIEGADVSTSKLLKAASPEKFTVA
 TDLGIHEMOKAPDKQFIAPAPAGNGSCSCAFPMAMNSLGGIKYALTSGRNEI
 ILDRKLEAKKLEIQRMLDPAAGLKKKDVNGMGA"
 6620. .7558
 /gene="NMB0395"
 6620. .7558
 /gene="NMB0395"
 /note="conserved hypothetical protein; identified by
 Glimmer2; putative"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA040835.1"
 /db_xref="GI:7225617"
 /translation="NDAYPEAEAPPOSIVELVPLIATVDGGLRVLTVAOGMLPNP
 LSPRLNSIAGVAKIYAKOTSOPMGVDEQIYTPVDHRRHEHMPVLYVYGLVREA
 ADSILHPDAKMDQCYGFPEWEDLRTDGGODAVYGRIRIANSADTBEVRORLKLREH
 LCGVEPEMNSSEYVQRYEMLYESGLIAAAPQANFDPALGQPMRHRVNLATA
 LSRLRAKIKRYPVIFELMPEFTLLQLONSVEAISRLHKNFRROIOOONLIEPDS
 TGVSGSKRPAQCRFRDVLPLRLISDIGLPLGSR"
 7776. .8657
 /gene="NMB0396"
 7776. .8657
 /gene="NMB0396"
 /note="similar to PID:665965 SP:P7938 PID:1498753 percent
 identity: 67.53; identified by sequence similarity;
 putative"
 /codon_start=1
 /transl_table=11
 /product="nicotinate-nucleotide pyrophosphorylase"
 /protein_id="AA040836.1"
 /db_xref="GI:7225618"
 /translation="MPESEKLPPLPTLLRPVYEALSEDLAGRDPITSAAVAPDKT
 AKFLVSRDEGVLAGMDIARLAFQTMDSRPAEIRDQAVAGQOTIAEVENARAL
 LAEPRTAINVTLHLSGIATATAVAVAEAYFGDYCSKRTPIILLVLOKAYRAGCG
 VNRKGLDPAVLKDNHLATCGIAQAVQAKAGVALTCVELEVDTLQDLDAIAG
 AERILLDNMDEDTLKEAANNCHTQTAHPHTIYCAASGIGFDRLKRVAOTGYVDGIALG
 YLTHSSRLSDIGDFAVA"
 8685. .8972

CDS
 /gene="NMB0397"
 8685. .8972
 /gene="NMB0397"
 /note="hypothetical protein; identified by Glimmer2;
 putative"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA040837.1"
 /db_xref="GI:7225619"
 /translation="MSGKEPLNPNPIYASRKRLKDDSGCIRKGRDALEKTAFOAG
 ILFARQGGIGRAGKRTGGSGFMSSEPLRLPAKRVGKTYPVSSINI"
 9064. .9345
 /gene="NMB0398"
 9064. .9345
 /gene="NMB0398"
 /note="similar to GB:000096 PID:1789021 PID:1800053
 PID:1800057 percent identity: 59.26; identified by
 sequence similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="transcriptional regulator, AtSR family"
 /protein_id="AA040838.1"
 /db_xref="GI:7225620"

Query Match 14.6%; Score 18; DB 1; Length 11018;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 97 ctatctcaattatc 114
 Db 10210 CTTATCTTCATTATTTC 10227

RESULT 16
 CELC34H3 33803 bp DNA INV 28-MAR-2000
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 REMARK

1
 1:283(5398):35 and 1999 Mar 26;283(5410):2103 and 1999 Sep
 3:286(5433):149311
 2 (bases 1 to 33803)
 Lamar, B. and Wamsley, P.
 The sequence of C. elegans cosmid C34H3
 Unpublished
 3 (bases 1 to 33803)
 Waterston, R.H.
 Direct Submission
 Submitted (23-FEB-1999) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 4 (bases 1 to 33803)
 Waterston, R.H.
 Direct Submission
 Submitted (01-MAR-2000) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA
 5 (bases 1 to 33803)

REFERENCE
 TITLE
 JOURNAL
 MEDLINE
 REMARK

AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (08-MAR-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE 6 (bases 1 to 33803)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (28-MAR-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT

Submitted by:

Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RQ, England
e-mail: twenematode.wustl.edu and jesusanger.ac.uk

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

NEIGHBORING COSMID INFORMATION:

The 5' clone is Y72A10A, 200 bp overlap; 3' clone is T25B2, 200 bp overlap. Actual start of this clone is at base position 1 of CELC34H3; actual end is at 33607 of CELC34H3

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFinder (P. Green and L. Hillier, ms in preparation).

FEATURES

Location/Qualifiers

1..33803

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="X"

/clone="C34H3"

/complement(8252..16090)

/gene="C34H3.1"

/complement(join(8252..8406,8464..8572,9494..9597,9653..9756,9809..9894,9946..10146,13178..13307,13542..13628,14243..14403,14887..14998,15046..15146,15195..15266,15855..15995,16043..16090))

/gene="C34H3.1"

/note="contains similarity to Pfam family PF00200 (Disintegrin), score=15.2, E=0.00052, N=1; coded for by C. elegans CDNA YK228a1.5; coded for by C. elegans CDNA YK344b12.5; coded for by C. elegans CDNA YK228a1.3; coded for by C. elegans CDNA YK344b12.3"

/codon_start=1

/protein_id="AAF39762.1"

/db_xref="GI:7206602"

/translation="MRKSLILVFCIFENNVSTLEOKHIEGAVGWAKADYSIAS

TDSGITVIRSLVFDNKTATYEDMIVKLNIMKAVDEANQYLNLQDLVGLVGL

QTRGDLISQSFHEYNRLKLPDHEFTLLSYRAGGLAVNGCSSHSVLSGFY

PNEPRAMGSIFFHEVAHLGVHRAVNESIYVNCICPEKDSKEGCKIKIGFDHC

TVOQEVNTIYKKNKILKHPILFEQSEVCGVLENDGDCGPGCGLNCPHPCR

FEMHPEFLVLAVAFVAVLVTWFTVRYGTGTMNCEPMYCKHGRGASPYTNGCI

QILASPYONRKMSSISGSWTILVNSRATIORKVPVPPPPPKRTIOVAVGS

GGQYFTTVFGSGYRESFYDDSDDEDEDESAIPPLPGVPCPAYPPNVPINRYTO

CLIPQNTHSCAPSSHQTFHFDAGAVLNIHMSILSFQGYFVODMLFSAHCFEFLYF

gene

LVLFFPNEFLKTTFTLMLIISPCR"

complement(25250..26090)

/gene="C34H3.2"

/complement(join(25250..25703,25780..26090))

/gene="C34H3.2"

/note="contains similarity to Pfam family PF00096 (zf-C2H2), score=77.4, E=3e-19, N=3"

/codon_start=1

/evidence="not_experimental"

/protein_id="AAF39763.1"

/db_xref="GI:7206603"

/translation="MLPMORVPTSIPOSNQVPRMMLAOLQLOLQORRMALL

AMNPEIPMTDLKAKKEDFTTHADSISSQKIKESVSPKSPITTTAAVPRPYDQ

PWMIPGRGRTTGKARPKKFTICKDHFPTSTNLLIHERTIDERPYSQDVCGRA

FRQDHLRDHKYIHOGRPKCEICGCGCOSRTLLVHRATHDPNRHSIGAVVPIKS

ETPLPELDPRVLLIQLNLDNSFNSTSMSPQISPDPR"

BASE COUNT

11637 a 5470 c 6333 g 10363 t

ORIGIN

Query Match

14.6%; Score 18; DB 3; Length 33803;

Best Local Similarity

100.0%; Pred. No. 18;

Matches

18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

99 tatctcaattattcca 116

Db

4201 TATCTCAATTATTCCA 4184

RESULT 17

LOCUS

AC022128 34128 bp DNA PRI 30-MAR-2001

DEFINITION

Homo sapiens chromosome 5 clone CTD-2170613, complete sequence.

ACCESSION

AC022128

VERSION

AC022128.5 GI:13489154

KEYWORDS

HTG.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS

Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

TITLE

DOE Joint Genome Institute and Stanford Human Genome Center.

JOURNAL

Direct Submission

REFERENCE

2 (bases 1 to 34128)

AUTHORS

DOE Joint Genome Institute.

TITLE

Direct Submission

JOURNAL

Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE

3 (bases 1 to 34128)

AUTHORS

DOE Joint Genome Institute and Stanford Human Genome Center.

TITLE

Direct Submission

JOURNAL

Submitted (30-MAR-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT

On Mar 30, 2001 this sequence version replaced gi:7711703.

REFERENCE

Draft Sequence Produced by DOE Joint Genome Institute

AUTHORS

www.jgi.doe.gov

TITLE

Finishing Completed at Stanford Human Genome Center

JOURNAL

www.shgc.stanford.edu

REFERENCE

Quality: Phrap Quality >=40 99.98% of Sequence;

AUTHORS

Estimated Total Number of Errors is 0.1.

TITLE

Note: Consensus clipped at overlap with AC008790.

JOURNAL

Location/Qualifiers

FEATURES

1..34128

source

/organism="Homo sapiens"

BASE COUNT

10725 a 7079 c 6751 g 9573 t

ORIGIN

/db_xref="taxon:9606"

Query Match

14.6%; Score 18; DB 9; Length 34128;

Best Local Similarity 100.0%; Pred. No. 18:
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 caacctgatyacctaat 45
|||||

Db 29094 CAACCTGATGACCTAAT 29111

RESULT 18

AC078862

57000 bp DNA

PRI

23-MAR-2001

LOCUS

Homo sapiens 12q BAC RP11-139B1 (Roswell Park Cancer Institute

DEFINITION

Human BAC library) complete sequence.

ACCESSION

AC078862

VERSION

AC078862.14 GI:13435185

KEYWORDS

HTG.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 57000)

AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks,S.L., Amaralungu,H.C., Are,D.R., Banks,T., Barbarta,J.,

Benton,D., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,

Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,

Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F.,

Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,

Chen,Z., Chiu,D., Chowdhury,I., Christopoulos,C., Cleveland,C.D.,

Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,

Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,

Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,

Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,

Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,

Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,

Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,

Gunnarath,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,

Hart,M., Haylak,P., Hawes,A., Hernandez,J., Hernandez,O.,

Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homsi,F.,

Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,

Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,

Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,

Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,

Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,

Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,

Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,

Marlinda,E., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P.,

Meador,M., Mel,G., Merscher,S., Metzker,M., Miller,A., Miner,G.,

Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,

Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,N.,

Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S.,

Ogun,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,

Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.,

Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubenkan,I., Rolfe,M.,

Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shm,C.,

Shooshitari,N., Sisson,I., Sodergren,E., Sonalke,T., Sparks,A.,

Stanley,R., Stone,H., Sutton,A., Svatek,A., Tabori,P., Tameris,A.,

Tameris,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B.,

Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalón,D.,

Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,

Washington,C., Watlington,S., Williams,G., Williamson,A.,

Walczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,

Zorilla,S., Kuchelapatti,R. and Gibbs,R.

JOURNAL COMMENT

Submitted (23-MAR-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Mar 23, 2001 this sequence version replaced gi:13173525.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

Contig length: 156673
Phrap values in estimate: 156142
Average error rate (BCM-Phrap estimate): 4.07525e-05
Fraction of Phrap values less than 40 : 0.00374019
Number of consensus changing edits: 306
Number of N's in consensus : 0

Position	Consensus changing edits	Edited+Context
36	aaaggtcctt(n)agtttcctc	aaaggtcctt(g)agtttcctc
6213	tctcttggtt(n)tcacagttt	tctcttggtt(c)tcacagttt
14543	tgcttcaccc(n)annnnnnnn	tgcttcaccc(t)agatggaat
14545	ctccatccca(n)atgnnnnnn	ctccatccca(g)atggaat
14546	lcccatccca(n)atgnnnnnn	lcccatccca(g)atggaat
14549	atcccaantg(n)nnnnnncc	atcccaantg(t)atggaat
14550	tcctcaantg(n)nnnnnncc	tcctcaantg(g)atggaat
14551	ccctcaantg(n)nnnnnncc	ccctcaantg(a)atggaat
14553	nnnnnnnn(n)nnnnnncc	nnnnnnnn(c)atggaat
14554	annnnnnnn(n)nnnnnncc	annnnnnnn(g)atggaat
14555	nnnnnnnn(n)nnnnnncc	nnnnnnnn(a)atggaat
14556	nnnnnnnn(n)nnnnnncc	nnnnnnnn(t)atggaat
14561	nnnnnnnn(n)nnnnnncc	nnnnnnnn(c)atggaat
14563	nnnnnnnn(n)nnnnnncc	nnnnnnnn(g)atggaat
14565	nnnnnnnn(n)nnnnnncc	nnnnnnnn(a)atggaat
14566	nnnnnnnn(n)nnnnnncc	nnnnnnnn(t)atggaat
14567	cccttgctta(n)nnnnnncc	cccttgctta(g)atggaat
14568	cccttgctta(n)nnnnnncc	cccttgctta(a)atggaat

```

14569 centtntnn(n)atnnncn
14572 tntnnnnat(n)nnncntn
14573 ntannnnat(n)nnncntn
14574 ntannnnat(n)nnncntn
14575 ntannnnat(n)nnncntn
14576 ntannnnat(n)nnncntn
14577 ntannnnat(n)nnncntn
14578 ntannnnat(n)nnncntn
14579 ntannnnat(n)nnncntn
14580 ntannnnat(n)nnncntn
14581 ntannnnat(n)nnncntn
14582 ntannnnat(n)nnncntn
14583 ntannnnat(n)nnncntn
14584 ntannnnat(n)nnncntn
14585 ntannnnat(n)nnncntn
14586 ntannnnat(n)nnncntn
14587 ntannnnat(n)nnncntn
14588 ntannnnat(n)nnncntn
14589 ntannnnat(n)nnncntn
14590 ntannnnat(n)nnncntn
14591 ntannnnat(n)nnncntn
14592 ntannnnat(n)nnncntn
14593 ntannnnat(n)nnncntn
14594 ntannnnat(n)nnncntn
14595 ntannnnat(n)nnncntn
14596 ntannnnat(n)nnncntn
14597 ntannnnat(n)nnncntn
14598 ntannnnat(n)nnncntn
14599 ntannnnat(n)nnncntn
14600 ntannnnat(n)nnncntn
14601 ntannnnat(n)nnncntn
14602 ntannnnat(n)nnncntn
14603 ntannnnat(n)nnncntn
14604 ntannnnat(n)nnncntn
14605 ntannnnat(n)nnncntn
14606 ntannnnat(n)nnncntn
14607 ntannnnat(n)nnncntn
14608 ntannnnat(n)nnncntn
14609 ntannnnat(n)nnncntn
14610 ntannnnat(n)nnncntn
14611 ntannnnat(n)nnncntn
14612 ntannnnat(n)nnncntn
14613 ntannnnat(n)nnncntn
14614 ntannnnat(n)nnncntn
14615 ntannnnat(n)nnncntn
14616 ntannnnat(n)nnncntn
14617 ntannnnat(n)nnncntn
14618 ntannnnat(n)nnncntn
14619 ntannnnat(n)nnncntn
14620 ntannnnat(n)nnncntn
14621 ntannnnat(n)nnncntn
14622 ntannnnat(n)nnncntn
14623 ntannnnat(n)nnncntn
14624 ntannnnat(n)nnncntn
14625 ntannnnat(n)nnncntn
14626 ntannnnat(n)nnncntn
14627 ntannnnat(n)nnncntn
14628 ntannnnat(n)nnncntn
14629 ntannnnat(n)nnncntn
14630 ntannnnat(n)nnncntn
14631 ntannnnat(n)nnncntn
14632 ntannnnat(n)nnncntn
14633 ntannnnat(n)nnncntn
14634 ntannnnat(n)nnncntn
14635 ntannnnat(n)nnncntn
14636 ntannnnat(n)nnncntn
14637 ntannnnat(n)nnncntn
14638 ntannnnat(n)nnncntn
14639 ntannnnat(n)nnncntn
14640 ntannnnat(n)nnncntn
14641 ntannnnat(n)nnncntn
14642 ntannnnat(n)nnncntn
14643 ntannnnat(n)nnncntn
14644 ntannnnat(n)nnncntn
14645 ntannnnat(n)nnncntn
14646 ntannnnat(n)nnncntn
14647 ntannnnat(n)nnncntn
14648 ntannnnat(n)nnncntn
14649 ntannnnat(n)nnncntn
14650 ntannnnat(n)nnncntn
14651 ntannnnat(n)nnncntn
14652 ntannnnat(n)nnncntn
14653 ntannnnat(n)nnncntn
14654 ntannnnat(n)nnncntn
14655 ntannnnat(n)nnncntn
14656 ntannnnat(n)nnncntn
14657 ntannnnat(n)nnncntn
14658 ntannnnat(n)nnncntn
14659 ntannnnat(n)nnncntn
14660 ntannnnat(n)nnncntn
14661 ntannnnat(n)nnncntn
14662 ntannnnat(n)nnncntn
14663 ntannnnat(n)nnncntn
14664 ntannnnat(n)nnncntn
14665 ntannnnat(n)nnncntn
14666 ntannnnat(n)nnncntn
14667 ntannnnat(n)nnncntn
14668 ntannnnat(n)nnncntn
14669 ntannnnat(n)nnncntn
14670 ntannnnat(n)nnncntn
14671 ntannnnat(n)nnncntn
14672 ntannnnat(n)nnncntn
14673 ntannnnat(n)nnncntn
14674 ntannnnat(n)nnncntn
14675 ntannnnat(n)nnncntn
14676 ntannnnat(n)nnncntn
14677 ntannnnat(n)nnncntn
14678 ntannnnat(n)nnncntn
14679 ntannnnat(n)nnncntn
14680 ntannnnat(n)nnncntn
14681 ntannnnat(n)nnncntn
14682 ntannnnat(n)nnncntn
14683 ntannnnat(n)nnncntn
14684 ntannnnat(n)nnncntn
14685 ntannnnat(n)nnncntn
14686 ntannnnat(n)nnncntn
14687 ntannnnat(n)nnncntn
14688 ntannnnat(n)nnncntn
14689 ntannnnat(n)nnncntn
14690 ntannnnat(n)nnncntn
14691 ntannnnat(n)nnncntn
14692 ntannnnat(n)nnncntn
14693 ntannnnat(n)nnncntn
14694 ntannnnat(n)nnncntn
14695 ntannnnat(n)nnncntn
14696 ntannnnat(n)nnncntn
14697 ntannnnat(n)nnncntn
14698 ntannnnat(n)nnncntn
14699 ntannnnat(n)nnncntn
14700 ntannnnat(n)nnncntn
14701 ntannnnat(n)nnncntn
14702 ntannnnat(n)nnncntn
14703 ntannnnat(n)nnncntn
14704 ntannnnat(n)nnncntn
14705 ntannnnat(n)nnncntn
14706 ntannnnat(n)nnncntn
14707 ntannnnat(n)nnncntn
14708 ntannnnat(n)nnncntn
14709 ntannnnat(n)nnncntn
14710 ntannnnat(n)nnncntn
14711 ntannnnat(n)nnncntn
14712 ntannnnat(n)nnncntn
14713 ntannnnat(n)nnncntn
14714 ntannnnat(n)nnncntn
14715 ntannnnat(n)nnncntn
14716 ntannnnat(n)nnncntn
14717 ntannnnat(n)nnncntn
14718 ntannnnat(n)nnncntn
14719 ntannnnat(n)nnncntn
14720 ntannnnat(n)nnncntn
14721 ntannnnat(n)nnncntn
14722 ntannnnat(n)nnncntn
14723 ntannnnat(n)nnncntn
14724 ntannnnat(n)nnncntn
14725 ntannnnat(n)nnncntn
14726 ntannnnat(n)nnncntn
14727 ntannnnat(n)nnncntn
14728 ntannnnat(n)nnncntn
14729 ntannnnat(n)nnncntn
14730 ntannnnat(n)nnncntn
14731 ntannnnat(n)nnncntn
14732 ntannnnat(n)nnncntn
14733 ntannnnat(n)nnncntn
14734 ntannnnat(n)nnncntn
14735 ntannnnat(n)nnncntn
14736 ntannnnat(n)nnncntn
14737 ntannnnat(n)nnncntn
14738 ntannnnat(n)nnncntn
14739 ntannnnat(n)nnncntn
14740 ntannnnat(n)nnncntn
14741 ntannnnat(n)nnncntn
14742 ntannnnat(n)nnncntn
14743 ntannnnat(n)nnncntn
14744 ntannnnat(n)nnncntn
14745 ntannnnat(n)nnncntn
14746 ntannnnat(n)nnncntn
14747 ntannnnat(n)nnncntn
14748 ntannnnat(n)nnncntn
14749 ntannnnat(n)nnncntn
14750 ntannnnat(n)nnncntn
14751 ntannnnat(n)nnncntn
14752 ntannnnat(n)nnncntn
14753 ntannnnat(n)nnncntn
14754 ntannnnat(n)nnncntn
14755 ntannnnat(n)nnncntn
14756 ntannnnat(n)nnncntn
14757 ntannnnat(n)nnncntn
14758 ntannnnat(n)nnncntn
14759 ntannnnat(n)nnncntn
14760 ntannnnat(n)nnncntn
14761 ntannnnat(n)nnncntn
14762 ntannnnat(n)nnncntn
14763 ntannnnat(n)nnncntn
14764 ntannnnat(n)nnncntn
14765 ntannnnat(n)nnncntn
14766 ntannnnat(n)nnncntn
14767 ntannnnat(n)nnncntn
14768 ntannnnat(n)nnncntn
14769 ntannnnat(n)nnncntn
14770 ntannnnat(n)nnncntn
14771 ntannnnat(n)nnncntn
14772 ntannnnat(n)nnncntn
14773 ntannnnat(n)nnncntn
14774 ntannnnat(n)nnncntn
14775 ntannnnat(n)nnncntn
14776 ntannnnat(n)nnncntn
14777 ntannnnat(n)nnncntn
14778 ntannnnat(n)nnncntn
14779 ntannnnat(n)nnncntn
14780 ntannnnat(n)nnncntn
14781 ntannnnat(n)nnncntn
14782 ntannnnat(n)nnncntn
14783 ntannnnat(n)nnncntn
14784 ntannnnat(n)nnncntn
14785 ntannnnat(n)nnncntn
14786 ntannnnat(n)nnncntn
14787 ntannnnat(n)nnncntn
14788 ntannnnat(n)nnncntn
14789 ntannnnat(n)nnncntn
14790 ntannnnat(n)nnncntn
14791 ntannnnat(n)nnncntn
14792 ntannnnat(n)nnncntn
14793 ntannnnat(n)nnncntn
14794 ntannnnat(n)nnncntn
14795 ntannnnat(n)nnncntn
14796 ntannnnat(n)nnncntn
14797 ntannnnat(n)nnncntn
14798 ntannnnat(n)nnncntn
14799 ntannnnat(n)nnncntn
14800 ntannnnat(n)nnncntn
14801 ntannnnat(n)nnncntn
14802 ntannnnat(n)nnncntn
14803 ntannnnat(n)nnncntn
14804 ntannnnat(n)nnncntn
14805 ntannnnat(n)nnncntn
14806 ntannnnat(n)nnncntn
14807 ntannnnat(n)nnncntn
14808 ntannnnat(n)nnncntn
14809 ntannnnat(n)nnncntn
14810 ntannnnat(n)nnncntn
14811 ntannnnat(n)nnncntn
14812 ntannnnat(n)nnncntn
14813 ntannnnat(n)nnncntn
14814 ntannnnat(n)nnncntn
14815 ntannnnat(n)nnncntn
14816 ntannnnat(n)nnncntn
14817 ntannnnat(n)nnncntn
14818 ntannnnat(n)nnncntn
14819 ntannnnat(n)nnncntn
14820 ntannnnat(n)nnncntn
14821 ntannnnat(n)nnncntn
14822 ntannnnat(n)nnncntn
14823 ntannnnat(n)nnncntn
14824 ntannnnat(n)nnncntn
14825 ntannnnat(n)nnncntn
14826 ntannnnat(n)nnncntn
14827 ntannnnat(n)nnncntn
14828 ntannnnat(n)nnncntn
14829 ntannnnat(n)nnncntn
14830 ntannnnat(n)nnncntn
14831 ntannnnat(n)nnncntn
14832 ntannnnat(n)nnncntn
14833 ntannnnat(n)nnncntn
14834 ntannnnat(n)nnncntn
14835 ntannnnat(n)nnncntn
14836 ntannnnat(n)nnncntn
14837 ntannnnat(n)nnncntn
14838 ntannnnat(n)nnncntn
14839 ntannnnat(n)nnncntn
14840 ntannnnat(n)nnncntn
14841 ntannnnat(n)nnncntn
14842 ntannnnat(n)nnncntn
14843 ntannnnat(n)nnncntn
14844 ntannnnat(n)nnncntn
14845 ntannnnat(n)nnncntn
14846 ntannnnat(n)nnncntn
14847 ntannnnat(n)nnncntn
14848 ntannnnat(n)nnncntn
14849 ntannnnat(n)nnncntn
14850 ntannnnat(n)nnncntn
14851 ntannnnat(n)nnncntn
14852 ntannnnat(n)nnncntn
14853 ntannnnat(n)nnncntn
14854 ntannnnat(n)nnncntn
14855 ntannnnat(n)nnncntn
14856 ntannnnat(n)nnncntn
14857 ntannnnat(n)nnncntn
14858 ntannnnat(n)nnncntn
14859 ntannnnat(n)nnncntn
14860 ntannnnat(n)nnncntn
14861 ntannnnat(n)nnncntn
14862 ntannnnat(n)nnncntn
14863 ntannnnat(n)nnncntn
14864 ntannnnat(n)nnncntn
14865 ntannnnat(n)nnncntn
14866 ntannnnat(n)nnncntn
14867 ntannnnat(n)nnncntn
14868 ntannnnat(n)nnncntn
14869 ntannnnat(n)nnncntn
14870 ntannnnat(n)nnncntn
14871 ntannnnat(n)nnncntn
14872 ntannnnat(n)nnncntn
14873 ntannnnat(n)nnncntn
14874 ntannnnat(n)nnncntn
14875 ntannnnat(n)nnncntn
14876 ntannnnat(n)nnncntn
14877 ntannnnat(n)nnncntn
14878 ntannnnat(n)nnncntn
14879 ntannnnat(n)nnncntn
14880 ntannnnat(n)nnncntn
14881 ntannnnat(n)nnncntn
14882 ntannnnat(n)nnncntn
14883 ntannnnat(n)nnncntn
14884 ntannnnat(n)nnncntn
14885 ntannnnat(n)nnncntn
14886 ntannnnat(n)nnncntn
14887 ntannnnat(n)nnncntn
14888 ntannnnat(n)nnncntn
14889 ntannnnat(n)nnncntn
14890 ntannnnat(n)nnncntn
14891 ntannnnat(n)nnncntn
14892 ntannnnat(n)nnncntn
14893 ntannnnat(n)nnncntn
14894 ntannnnat(n)nnncntn
14895 ntannnnat(n)nnncntn
14896 ntannnnat(n)nnncntn
14897 ntannnnat(n)nnncntn
14898 ntannnnat(n)nnncntn
14899 ntannnnat(n)nnncntn
14900 ntannnnat(n)nnncntn
14901 ntannnnat(n)nnncntn
14902 ntannnnat(n)nnncntn
14903 ntannnnat(n)nnncntn
14904 ntannnnat(n)nnncntn
14905 ntannnnat(n)nnncntn
14906 ntannnnat(n)nnncntn
14907 ntannnnat(n)nnncntn
14908 ntannnnat(n)nnncntn
14909 ntannnnat(n)nnncntn
14910 ntannnnat(n)nnncntn
14911 ntannnnat(n)nnncntn
14912 ntannnnat(n)nnncntn
14913 ntannnnat(n)nnncntn
14914 ntannnnat(n)nnncntn
14915 ntannnnat(n)nnncntn
14916 ntannnnat(n)nnncntn
14917 ntannnnat(n)nnncntn
14918 ntannnnat(n)nnncntn
14919 ntannnnat(n)nnncntn
14920 ntannnnat(n)nnncntn
14921 ntannnnat(n)nnncntn
14922 ntannnnat(n)nnncntn
14923 ntannnnat(n)nnncntn
14924 ntannnnat(n)nnncntn
14925 ntannnnat(n)nnncntn
14926 ntannnnat(n)nnncntn
14927 ntannnnat(n)nnncntn
14928 ntannnnat(n)nnncntn
14929 ntannnnat(n)nnncntn
14930 ntannnnat(n)nnncntn
14931 ntannnnat(n)nnncntn
14932 ntannnnat(n)nnncntn
14933 ntannnnat(n)nnncntn
14934 ntannnnat(n)nnncntn
14935 ntannnnat(n)nnncntn
14936 ntannnnat(n)nnncntn
14937 ntannnnat(n)nnncntn
14938 ntannnnat(n)nnncntn
14939 ntannnnat(n)nnncntn
14940 ntannnnat(n)nnncntn
14941 ntannnnat(n)nnncntn
14942 ntannnnat(n)nnncntn
14943 ntannnnat(n)nnncntn
14944 ntannnnat(n)nnncntn
14945 ntannnnat(n)nnncntn
14946 ntannnnat(n)nnncntn
14947 ntannnnat(n)nnncntn
14948 ntannnnat(n)nnncntn
14949 ntannnnat(n)nnncntn
14950 ntannnnat(n)nnncntn
14951 ntannnnat(n)nnncntn
14952 ntannnnat(n)nnncntn
14953 ntannnnat(n)nnncntn
14954 ntannnnat(n)nnncntn
14955 ntannnnat(n)nnncntn
14956 ntannnnat(n)nnncntn
14957 ntannnnat(n)nnncntn
14958 ntannnnat(n)nnncntn
14959 ntannnnat(n)nnncntn
14960 ntannnnat(n)nnncntn
14961 ntannnnat(n)nnncntn
14962 ntannnnat(n)nnncntn
14963 ntannnnat(n)nnncntn
14964 ntannnnat(n)nnncntn
14965 ntannnnat(n)nnncntn
14966 ntannnnat(n)nnncntn
14967 ntannnnat(n)nnncntn
14968 ntannnnat(n)nnncntn
14969 ntannnnat(n)nnncntn
14970 ntannnnat(n)nnncntn
14971 ntannnnat(n)nnncntn
14972 ntannnnat(n)nnncntn
14973 ntannnnat(n)nnncntn
14974 ntannnnat(n)nnncntn
14975 ntannnnat(n)nnncntn
14976 ntannnnat(n)nnncntn
14977 ntannnnat(n)nnncntn
14978 ntannnnat(n)nnncntn
14979 ntannnnat(n)nnncntn
14980 ntannnnat(n)nnncntn
14981 ntannnnat(n)nnncntn
14982 ntannnnat(n)nnncntn
14983 ntannnnat(n)nnncntn
14984 ntannnnat(n)nnncntn
14985 ntannnnat(n)nnncntn
14986 ntannnnat(n)nnncntn
14987 ntannnnat(n)nnncntn
14988 ntannnnat(n)nnncntn
14989 ntannnnat(n)nnncntn
14990 ntannnnat(n)nnncntn
14991 ntannnnat(n)nnncntn
14992 ntannnnat(n)nnncntn
14993 ntannnnat(n)nnncntn
14994 ntannnnat(n)nnncntn
14995 ntannnnat(n)nnncntn
14996 ntannnnat(n)nnncntn
14997 ntannnnat(n)nnncntn
14998 ntannnnat(n)nnncntn
14999 ntannnnat(n)nnncntn
15000 ntannnnat(n)nnncntn

```

Query Match 14.6% Score 18; DB 9; Length 57000;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 106 aattattccacacaca 123
 |||||
 Db 50562 AATTATTCCACACACA 50579

RESULT 19
 T8K14
 LOCUS
 DEFINITION
 Arabidopsis thaliana chromosome 1 BAC T8K14 sequence, complete
 sequence.
 ACCESSION
 NC007202
 VERSION
 GI:12039264
 HTG.
 SOURCE
 thale cress.
 Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 1 (bases 1 to 80374)
 Vysotskaya,V.S., Schwartz,J.R., Yu,G., Toriumi,M., Lenz,C., Liu,S.,
 Li,J., Kremenetskaia,I., Lunos,J., Lee,J.M., Gonzalez,A.,
 Alfati,H., Araujo,R., Chao,Q., Conn,L., Conway,A.B., Dunn,P.,
 Hansen,N., Huizar,L., Kim,C., Palm,C., Rowley,D., Shinn,P.,
 Walker,M., Davis,R.W., Ecker,J.R., Federspiel,N.A. and Theologis,A.
 Arabidopsis thaliana chromosome 1 BAC T8K14 sequence
 Unpublished
 2 (bases 1 to 80374)
 Theologis,A.
 Direct Submission
 Submitted (03-APR-1999) Plant Gene Expression Center, 800 Buchanan
 Street, Albany, CA 94710, USA
 3 (bases 1 to 80374)
 Theologis,A.
 Direct Submission
 Submitted (29-APR-1999) Plant Gene Expression Center, 800 Buchanan
 Street, Albany, CA 94710, USA
 4 (bases 1 to 80374)
 Theologis,A.
 Direct Submission
 Submitted (15-MAY-1999) Plant Gene Expression Center, 800 Buchanan
 Street, Albany, CA 94710, USA
 5 (bases 1 to 80374)
 Theologis,A.
 Direct Submission
 Submitted (05-JAN-2001) Plant Gene Expression Center, 800 Buchanan
 Street, Albany, CA 94710, USA
 COMMENT
 On Jan 5, 2001 this sequence version replaced gi:471943.
 This sequence is of BAC T8K14 from Arabidopsis thaliana chromosome
 1. In order to facilitate the joining of overlapping clones in the
 future for creation of larger contigs, we provide overlap between
 overlapping submitted clones. The 3' end of this sequence overlaps
 by 2000 bp the 5' end of the sequence of the YAC YUP8H12R.
 FEATURES
 source
 1. 80374
 /organism="Arabidopsis thaliana"
 /cultivar="Columbia"
 /db_xref="taxon:3702"
 /chromosome="1"
 /clone="T8K14"
 48. 5132
 /gene="T8K14.1"
 /join(448..3246,3320..3391,3652..3804,3894..4057,
 4156..4303,4397..4519,4737..4824,4933..5132)
 /gene="T8K14.1"
 /note="Is a member of the P100069 Eukaryotic protein
 kinase family. ESTs gb|T46484, gb|AF066875 and gb|N96237
 come from this gene."
 /codon_start=1
 /evidence-not-experimental
 /protein_id="AAD30219.1"
 /db_xref="GI:483752"
 /translation="WDKARHQLFQHSMEPEYRNETVPEPMDQTSASANKPPNS
 NGSVDKAVHNFSTIGEEFSLFMRDRVTPORSSPNMGADPMNTGMYELRGIGIS
 HFGSECAVDAREFVENGSTDIERTNSLSHEFGNKLNHVOSAPQALSDSSGNLH
 GYKNSSSASGVAKYKILISPGSKILPRGDSKILRYGEMPIISIRDDISMOERL
 OKLEIYQTRVYKQLPGEGLDLVSVSSSEDDLNMLEYENENENSGSKLPMFLF
 SISMDADLGVNKNRNDGSEFQYVAVNGMDISGKNSSTLLGLDSSSNLAELDVFN

TEGINTIAGDVGVGASOLMVMNGFOOTSACQSESIPISSSLHYSSOSIPLNAAVOLOS
VPSRALHYPOSTIPGSSLOYPOSITPGSSSYOYPOSTIPGASSYGIYPOYVHVHVOH
GERBEPRLYVPHSSSNYSATIGETTSISDIPGCHYOQOOGMAGCYPRPGSTPSTQALAE
OKVSSDMKIREVEPENRKTIPGNDHQPPOIDVEVNNQVREMAVATPPSODALH
LPPSRDPRONTAKPATYRDVAITGVPLSGIDQLSTSSSTYAPVHSDSNLIDN
YPPREQSSQRYVCSEIRIPROELNLNLSKSDNSLQSTSSSTYAPVHSDSNLIDN
GKSHDEKTYVNDANHHHTKQVETIFEKVGSDETLESEPLHKIVNPDANKRVVNG
ADREIGSYNLSHVNAAMSHVTPPEQASLOGDILIDINDRPPDPLSEIFSOALSEDT
TVRYPHSDGAASVMNVONHDKRMKSYTQQLAEQOIFORDVYLDQASRIPSDKDGSE
SRPLPYLSRDSGITNLNANPOLITGQDTGNGSEKGGGTSIPALNEQKATE
SEEFGAVENTLRTPDSEPKDEKTETRAALPLGSEFDYSGLOIKNEDLELELGS
GTFGTGYHGMKWSDYAIRIKKSCFGESEDEBRLGEPWGEAEILISKJHNNVAVF
YGVKDGPGCGTLATVEYVYDGLRHVLVDRDLDRKRLIITAMDAFEMKRLHNSN
TVHEDLKCDNLVLNKPDRPICKVGDGFLSKIKRMTLVGSVGRITLPMMAPELHNS
SSRYSEKVDVPSFGIYVIMELTGSEEPANMHGAILTGIGVNNLPRITIPFCDDERT
LMEECANPMAPSTETIAGRLRMSSAATSTQSKPSAHRASK"

gene
complement(5680..11012)
/gene="TRK14.2"
complement(join(5680..5757,5848..5928,6021..6128,
6221..6295,6397..6516,6700..6876,6987..7103,7197..7316,
7406..7443,7567..7787,7887..7999,8095..8262,8367..8507,
8611..8853,8933..9010,9243..9520,9737..9963,10170..10354,
10584..11012))
/note="Is a member of PFI0004 ATPases associated with
various cellular activities (AAA) family. ESTs gb|TA3031,
gb|R64750, gb|AA394742 and gb|AI100347 come from this
gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD30220.1"
/db_xref="GI:4835753"

CDS
/translation="MEIAISYKPNLTISSSTOLIKRSKSGIVRPKAYGATRRKO
LRFVYASSSSSSSNDGFSWNRILAQSTRIGAEKIGESVKTETPFDESEASG
KVMDEYARVDSVHKGHLELFRKNETVPSFIDMKNMEHMKDIRMWDGKRVAALETFA
FALLSCQRYVAIIQAPVERERELTSEMEALIPESPNGINKEKERNMRAKATPQD
LKLRFEADPGTLVHDSYVGENAMDDLETTEGSLKIGRNARIOTPAKKKLSQD
LGVSEIGDSVGNMRELRATWREKLEKLESEOLNSSAAKYVEFMDEKESLSEYV
IGRSTEGTRALMISKRMWRKLPYTYFLOKIDSEVAAYVEFEDLRIVTME
GPLEYIVDIPDIPYLETICNAGVEVDLLQKROIHEKMFYALDPLGITLIMFIRS
AMLLITSRKFLTKTKNOLEDMAYAEMLIPGVDSFTSKMKEVYLGVSAGWDLDEL
MIYGNPMQYKEDVAVRGVLLSGPRTGKTLEFARTLAKESGILPVGASGAETDSE
KSGAAKINEMESIARRNAPAFVEDEIDAIAGRRARPRRRATFEALIDLGEEKK
TGIDRESLRQAVFICATNRPDELEFVSGRDRLYIGLPDOKOROIOFGVHSG
KNLAEDIDECGANIRNLVNEAIIMSYKRGYIYODIYVLDKOLLEGVALLTEEE
QCKCEOSVSEYKRRILAVHAGHIVLAHLFPRPDMAHFSOLLPGKGTANSVYPRD
WVDQGTTFETKMQNVAVHAGRCQAEKVYVGDVTDGKDLERIKTKIAREVITSPDS
ARLGLTQLVKRIQMVDLPNDGELIKYRMDHGVMAEMSVSELSFTRELRYIEE
TEELAMNALLRANHIDLLITRELLEKSRITGLVEEEMKDLSPLMEDFVKPFQINDP
DEELPHKDRSVOPVDLRAAPLHRS"

gene
11693..13641
/gene="TRK14.3"
join(11693..11887,12188..12265,12354..12614,12703..13020,
13129..13287,13447..13641)
/note="Is a member of the PFI00162 Phosphoglycerate kinase
family. ESTs gb|N38721, gb|T22178, gb|P90345, gb|P90715,
gb|T21440, gb|T46295, gb|H37082, gb|T46076, gb|N37132,
gb|AA597649, gb|AI100648 and gb|T48462 come from this
gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD30221.1"
/db_xref="GI:4835754"

CDS
/translation="MAIKRSVGTGLKADLKSGSVFVRVDLNVPLDONSITDDTRIRA
AVPTIKLMGNSRVVLCSHLGRPKGVTPYSLSKPLVPRSELGVEVNAVANDSIGE
VOKLVAGLPEGVILLENVFAVEEKNDEFAKLLAALADVVDNDFGTAHRAHAST
EGVAKFLKPSVAGELMOKELDYLVGAVANKRPFAIVGSKSVSTKIGVETSLNPD
IILGGMITFEKKAGLSVSSVLEDKDLKLSMEKKAAGVSIILPTDYIADK
FAPDANSKIYPAITAPDGMGLDIDGPSIKTFSEADLTITIIWNGMGVFEEDKRP
GTEAVAKQLAELSGKVYTIIGGDSVAAEKVLADKMSHISTGGASLELLEGRPL
PGVIALDEA"

gene
15312..17654

CDS
/gene="TRK14.4"
15312..17654
/note="Contains similarity to g112827663 F18F4.190
membrane-associated salt-inducible-like protein from
Arabidopsis thaliana BAC gb|AU021637."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD30222.1"
/db_xref="GI:4835755"

gene
/translation="MKPQMFVSYIOFYSKPSWMSYSSGNAEFNISGEVISILAKK
KPIEPALEVPPLFSKNIITSVSKIDENRQDGFVFWAASRELRBSRESEGLVIDL
SENGCGLQVOTLEELKSGSVSDSYSCVYLISAYAMGMAERLBSFGMKEDCDP
DVEFYNNILVMMREEVFPMIAVYVEMKCNCPULYFEGILMDGLYKGRFTSDP
KMPDMMGRIISPNRYVYITLIGICQSGADDAARKIFYEMQTSNGYDPSVANAALD
GCKLGMVAFAELRLFEEDGYVLGRGSSILDGLFRARRITQAEIYANMLKKNII
KPIIITVTLIIQISKAGKIEDALKLSMPSKGISPDYCYANAVIALCGRLLEG
RSLOLEMSFESPEPDACTHTILICSMCRNLVRAEIEFIEIKSGSPVAFFNALI
DGCKSGEILKEARLLHKMEVGRPASLELHSHSGNSFDTWESGSLKAYRDLAHF
ADGSSPDIVSYVYLIFGCRAGIDGALKLVNLQKLSGSPVYKNTLNLGLHRVG
REERAEFLYAKNDRSPAVYSISMSCRKRVYVAFNLMKMYLTKISCLIDETAN
EIEOCREGETERARLRIELDRKDELITGPTTITLIGCQSRFHEALMVSYLEE
KILVTPPSCVKLIHIGCKREQLDALEVLTYLDNNFKLMPRCVNLISLLESTEK
MEIVSQLTNMRERAGVNDSEMLREILKYHRHROYLDL"

gene
18900..21756
/gene="TRK14.5"
join(18900..18995,19434..19507,19606..19673,19753..19791,
19928..20028,20168..20280,20369..20468,20547..20693,
20774..20834,20925..21022,21110..21252,21361..21444,
21529..21618,21702..21756)
/note="Is a member of the PFI00044 glyceraldehyde
3-phosphate dehydrogenase family. ESTs gb|T43985,
gb|N38667, gb|N65037, gb|AA713069 and gb|AI109548 come
from this gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD30223.1"
/db_xref="GI:4835756"

gene
/translation="MARSSILRSASVTVAAPDPFSSPASDHSKYLSISIGPSRLK
PSRSTISSSLONGNARSVQPIKATVATYPSAVRNSSSGKTVGNGRIGRLVYL
RIATSRDIEVVAANDPFIQAKYATVYLYKDSHNGKGSINVIDSLENGRILV
VSKRDPSEIEMADIGADYVVESSGVFTLLKASHLGGAKKVYISAPSDAPFVVG
VNEHTYQPNMDIYVNASCTNCLPLAKVHAEESGLEMTYHATTAQKYDGS
MKMRBGRGSRGRIIPSTGQAKAVAGVLELNKLGMAFRVTSVAVYDLCRIE
KGASVDEYKAIKHAISGPIKGLIGTYDEDEVYSDPFGDSSSIFDANAIGLSKSY
KLVSWIDNENGISNRYLDLIEHMLVAASH"

gene
23473..26525
/gene="TRK14.6"
join(23473..23665,24252..24364,24454..24699,24777..24857,
24955..25185,25275..25613,26424..26525)
/note="EST gb|AA404917 comes from this gene."
/codon_start=1
/evidence=not_experimental
/protein_id="AAD30224.1"
/db_xref="GI:4835757"

Query Match 14.6% Score 18; DB 8; Length 80374;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 gtatgttaatgttagat 21
|||||
Db 28318 GTATTGTAAGTGAGAT 28335
|||||

RESULT 20
AC008790/c AC008790 117018 bp DNA 09-SEP-2000
LOCUS Homo sapiens chromosome 5 clone CTD-2040H20, complete sequence.
DEFINITION AC008790
ACCESSION AC008790
VERSION AC008790.6 GI:10044330

gene
join(25341.25842,28143.28218,28300.28415,28495.29009)
/gene="P0445D12.4"
join(25341.25842,28143.28218,28300.28415,28495.29009)
/gene="P0445D12.4"
/note="contains ESTs A0161457(C63275),A0161458(C63275)
unknown protein"
/codon_start=1
/protein_id="BAB40033.1"
/db_xref="GI:13486801"
/translation="MGAADKWLPLVSVSVLLELALSGFSAASSLFLRPPSY
VRRGAAPSPFAVLSGREGKRLRLAVYHPRNRYLHLSDAPSERVELAA
VSRAPAVAFAGVDVAGPTAGTIMGSGIAVTLRAAALLRLDSEMPFTLNAD
YPLVTDODLHVSVSRHLNFDHTSDIGMKETORVPIYDAGIYIAGRNQFOAT
EKRDTPDGFEFFGSPWILNRPETICJGEMNLDPRTLLMTYNMLDQGFHSYV
CNSDFRSTVNSDMRYEMNDPPQMEPHFLNTHYIEIVSGVFPARKRENPDLIDK
IDRVLHRRHRRPVGAWCTGKRKRMFNDPCSNVNIYVPGFAERFRHMOIIE
SASGNSCKQ"
join(33714.33962,34268.35005)
/gene="P0445D12.5"
join(33714.33962,34268.35005)
/gene="P0445D12.5"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB40034.1"
/db_xref="GI:13486802"
/translation="MVEONKRRARNGEVOQOEEFEVEGEVSOEFTORTGAPL
VYMAAMEQTEETQIDVRIAVALLHCHACLOPKRPYFCDEHIVCSGRCRHHOL
CGAAVYSHCAELDAIVATAKPCAHAPGCSYVYAGADHQRACPCAPCCEPG
CRRPSPALPGHLAGHSMPVAIEYGRKRAVPPRHVLGEADRVFLVLSGAV
GAGAACVVCVRANGGDNAAVAARYCKLWEPNDNMAMTSMVSSDLGSGFP
AADKGLWVPEMLHGVPGGEAAILSTRDRAAATPFTTTRAFSOGMH"
complement(join(38187.38413,38479.38682,39059.39788))
/gene="P0445D12.6"
complement(join(38187.38413,38479.38682,39059.39788))
/gene="P0445D12.6"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB40035.1"
/db_xref="GI:13486803"
/translation="MGRPARREGPRHARGRGASGSPRVVPSGTRLSVARTVR
REHARARIRMGDAVHARGSRGCGGPTRIAMHAGPRARAGORAAADRGDPLA
ELAPTRWRGCHGTGREVEDAGONGRTAATGAGNHGDTGGSGTGMHGERDPA
TARIRRELDDGGILRRROPASRGNGDCAQREPRASPTLRSSVSRLDQPT
REAGDYMALPTSGDGCETASGNGRSKTLALGCHONGREDDDAGCEKEERGE
RRLASCHGFRKRREREGDAVGGALPILGSRGQALTAATGRSATTARAGS
NGAGDSKLGSGARERDEGEQENKEGELSGASAHARATHGDEEG"
join(42735.42995,43154.43891)
/gene="P0445D12.7"
join(42735.42995,43154.43891)
/gene="P0445D12.7"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB40036.1"
/db_xref="GI:13486804"
/translation="MAEONKRRSSAAENHAGCKRARAVEVKQEEFEADEGMSOE
ESAKTPIVATRAMDPQIDRIANGCLHCHACLOPKRPVCEAHAVSCCRNH
GQLCRRAAVYAHCAELDAIVAKVACAHPGCDYVYVGAADHQRACPCAPCCEP
DPCCGGRSPALIGHFADHPMSVTOISYAKRCRLAVPLPRCHVLGEEDRAMPLV
VSPSPCGVGAACVAVRANGDDAAAOYKCKLWEPNTSMVSSDLGSGFP
SGGFPAAEGMFLVVPPELHVEVSGEPTLSIRIDRAAPAIKPTTPRARSRLRLQ"
complement(join(44730.44939,45033.45272))
/gene="P0445D12.8"
complement(join(44730.44939,45033.45272))
/gene="P0445D12.8"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB40037.1"
/db_xref="GI:13486805"
/translation="MARIAMTVRRANOGRAVGDTHMAHSSRGSDOREGGVAGTACGD
RAACGRPARADAGHSCSGSGCGATGATVATRRPATPPRIVASVADMAAPCSRSDGP
PRMWSGRRSGRQPCGVRAARDYVHLCRSGRRDGGATAGVAVRH"

gene
join(46910.47188,47279.48013)
/gene="P0445D12.9"
join(46910.47188,47279.48013)
/gene="P0445D12.9"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB40038.1"
/db_xref="GI:13486806"
/translation="MONGEGSGAKRAAWVSPNGOVKREMAVEAARGESAAAAGS
EEREYVAGGIMTAAYAGDEGEYISVRIDLVLHCPILCLPLKPPYOCAGHLACS
SCHGDVPGKCHTCGGGGGCGVYARCPGLDTFLRAKTLIDPNLFCRSTVATHVAA
HORACHPACSCSEPRCDLFGSPMLLHLVADHSPVSVKPYGEVLJTJHVPSEERR
LVVAGAGGDEDERVFLVSGALGVAVASVACVRAANAAGPRYRCRLMHAAPCGAAD
FVMDASVAVSAAPGCEVAVDEARFLVPPCFLLDLAAGTSKEMLIRLSISIDIDI
DMS"
complement(join(48421.48594,48714.48845))
/gene="P0445D12.10"
complement(join(48421.48594,48714.48845))
/gene="P0445D12.10"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB40039.1"
/db_xref="GI:13486807"
/translation="MGBRWPAATEAVAKKVAAGVGLVAPDNMARCPACEDDDA
PGPADLRLRNFGNLTETIDKRLPTKRSEKPPYHFRYHRRKRLPPFPPIITDID"
complement(50655.51095)
/note="3' UTR"
complement(51250.54757)
misc_feature
LTR
Query Match 14.6% Score 18; DB 8; Length 140952;
Best local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 80 tccatgttaagtgatc 97
|||||
DB 62998 TCACATGTAAGTGATAC 63015
RESULT 22
AL161454/c
LOCUS
DEFINITION
Human DNA sequence from clone RP11-7284 on chromosome 9, complete
sequence.
ACCESSION
AL161454
VERSION
AL161454.10 GI:14456186
KEYWORDS
HTG.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 144536)
REFERENCE
AUTHORS
Williams, S.
TITLE
Direct Submision
JOURNAL
Submitted (13-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT
On Jun 14, 2001 this sequence version replaced gi:14269905.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP

database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Ch9>
RP11-72B4 is from the library RP11-11.1 constructed by the group of Plietser de Jong. For further details see
<http://www.chori.org/Dacpac/home.htm>
VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-72B4. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.
The true left end of clone RP11-72B4 is at 1 in this sequence. The true left end of clone RP11-18014 is at 142537 in this sequence. The true right end of clone RP11-346B7 is at 81734 in this sequence.

FEATURES

Source	Location/Qualifiers
repeat_region	1..144536
repeat_region	/organism="Homo sapiens"
repeat_region	/db_xref="taxon:9606"
repeat_region	/chromosome="9"
repeat_region	/clone="RP11-72B4"
repeat_region	/clone_lib="RP11-11.1"
repeat_region	1935..2764
repeat_region	/note="L2 repeat: matches 1852..2745 of consensus"
repeat_region	3353..3419
repeat_region	/note="Alu repeat: matches 233..298 of consensus"
repeat_region	4106..4149
repeat_region	/note="MIR repeat: matches 103..146 of consensus"
repeat_region	5694..5993
repeat_region	/note="Alusx repeat: matches 1..302 of consensus"
repeat_region	6069..6317
repeat_region	/note="MIR repeat: matches 15..262 of consensus"
repeat_region	7021..7128
repeat_region	/note="MIR repeat: matches 15..262 of consensus"
repeat_region	7130..7864
repeat_region	/note="MIR repeat: matches 1..184 of consensus"
repeat_region	8159..8304
repeat_region	/note="L1MC4 repeat: matches 6822..7608 of consensus"
repeat_region	8324..8359
repeat_region	/note="L1MC4 repeat: matches 7813..7952 of consensus"
repeat_region	8622..8704
repeat_region	/note="18 copies 2 mer ac 97% conserved"
repeat_region	8702..8757
repeat_region	/note="MER47 repeat: matches 2242..2323 of consensus"
repeat_region	9384..9430
repeat_region	/note="MER47 repeat: matches 2265..2320 of consensus"
repeat_region	9431..9784
repeat_region	/note="L1M4 repeat: matches 5276..5321 of consensus"
repeat_region	9785..9897
repeat_region	/note="THE1B repeat: matches 1..364 of consensus"
repeat_region	9948..10382
repeat_region	/note="L1M4 repeat: matches 5160..5276 of consensus"
repeat_region	10411..10623
repeat_region	/note="MTRC repeat: matches 20..466 of consensus"
repeat_region	10651..10777
repeat_region	/note="L1M4 repeat: matches 3504..3744 of consensus"
repeat_region	10794..11175
repeat_region	/note="Alub repeat: matches 84..206 of consensus"
repeat_region	11735..12861
repeat_region	/note="MTRD repeat: matches 1..417 of consensus"
repeat_region	12854..13614
repeat_region	/note="L1PA3 repeat: matches 4256..5387 of consensus"
repeat_region	14647..15016
repeat_region	/note="L1PA3 repeat: matches 5359..6142 of consensus"
repeat_region	15137..15290
repeat_region	/note="THE1C repeat: matches 1..371 of consensus"
repeat_region	15915..16173
repeat_region	/note="MERSA repeat: matches 9..178 of consensus"
repeat_region	17298..17542
repeat_region	/note="L1MC2 repeat: matches 5913..6162 of consensus"
repeat_region	17480..17590
repeat_region	/note="MIR repeat: matches 15..262 of consensus"
repeat_region	17729..17848
repeat_region	/note="Single clone region. Assembly confirmed by restriction digest data."
repeat_region	17915..18209
repeat_region	/note="60 copies 2 mer tt 60% conserved"
repeat_region	18257..18527
repeat_region	/note="Alu repeat: matches 1..296 of consensus"
repeat_region	18526..18645
repeat_region	/note="Alu repeat: matches 5893..6175 of consensus"
repeat_region	18647..18889
repeat_region	/note="L1 repeat: matches 3985..4104 of consensus"
repeat_region	18915..19315
repeat_region	/note="Alusx repeat: matches 64..302 of consensus"
repeat_region	19316..19616
repeat_region	/note="L1 repeat: matches 3596..3984 of consensus"
repeat_region	19617..21010
repeat_region	/note="Alusx repeat: matches 1..301 of consensus"
repeat_region	21041..21231
repeat_region	/note="L1 repeat: matches 2138..3596 of consensus"
repeat_region	21275..21472
repeat_region	/note="L1 repeat: matches 1910..2104 of consensus"
repeat_region	21473..21774
repeat_region	/note="L1MC4 repeat: matches 1489..1687 of consensus"
repeat_region	21775..22900
repeat_region	/note="Alus81 repeat: matches 3..296 of consensus"
repeat_region	22902..23133
repeat_region	/note="L1MC4 repeat: matches 844..1489 of consensus"
repeat_region	23135..24875
repeat_region	/note="Alusx repeat: matches 1..239 of consensus"
repeat_region	24965..25257
repeat_region	/note="TIGER1 repeat: matches 1..1792 of consensus"
repeat_region	25260..25323
repeat_region	/note="Alu repeat: matches 1..293 of consensus"
repeat_region	25777..26864
repeat_region	/note="L1M4 repeat: matches 2510..2574 of consensus"
repeat_region	26961..27967
repeat_region	/note="MER1C repeat: matches 1..1071 of consensus"
repeat_region	27974..28088
repeat_region	/note="L1M3 repeat: matches 5173..6178 of consensus"
repeat_region	30417..30572
repeat_region	/note="MERSA repeat: matches 46..159 of consensus"
repeat_region	30647..30901
repeat_region	/note="L1M4 repeat: matches 21..189 of consensus"
repeat_region	30898..31021
repeat_region	/note="L1M4 repeat: matches 2552..2809 of consensus"
repeat_region	31061..32002
repeat_region	/note="L1M4 repeat: matches 2773..2896 of consensus"
repeat_region	32004..32557
repeat_region	/note="L1M4 repeat: matches 179..1438 of consensus"
repeat_region	32601..34100
repeat_region	/note="L1MC repeat: matches 1761..2326 of consensus"
repeat_region	34099..34957
repeat_region	/note="L1MC repeat: matches 2406..3590 of consensus"
repeat_region	37554..37636
repeat_region	/note="L1M1 repeat: matches 5374..6224 of consensus"
repeat_region	37637..37941
repeat_region	/note="MIR repeat: matches 177..256 of consensus"
repeat_region	37942..38084
repeat_region	/note="Alusx repeat: matches 1..304 of consensus"
repeat_region	38862..38960
repeat_region	/note="MIR repeat: matches 22..177 of consensus"
repeat_region	39151..39429
repeat_region	/note="L1M4 repeat: matches 4487..4585 of consensus"
repeat_region	39570..40180
repeat_region	/note="Alusx repeat: matches 11..294 of consensus"
repeat_region	40736..40792
repeat_region	/note="HAL1 repeat: matches 414..1081 of consensus"
repeat_region	44656..45707
repeat_region	/note="HAL1 repeat: matches 883..939 of consensus"
repeat_region	50430..50717
repeat_region	/note="L1MC2 repeat: matches 5193..6224 of consensus"
repeat_region	50912..52140
repeat_region	/note="AlusC repeat: matches 1..287 of consensus"

```
repeat_region /note="TIGER2 repeat: matches 1461. .2718 of consensus"
52344. .52629
/note="Alusg repeat: matches 1. .291 of consensus"
repeat_region 54520. .54819
/note="Alusg repeat: matches 1. .298 of consensus"
repeat_region 56188. .56301
/note="L1MC4 repeat: matches 6570. .6692 of consensus"
repeat_region 56527. .56852
/note="L1MB6 repeat: matches 4937. .5258 of consensus"
repeat_region 56853. .57142
/note="Alusg repeat: matches 1. .290 of consensus"
repeat_region 57143. .57847
/note="L1MB6 repeat: matches 5258. .6004 of consensus"
repeat_region 57861. .57893
/note="MER3 repeat: matches 17. .49 of consensus"
repeat_region 58211. .58606
/note="L1MC4 repeat: matches 7031. .7419 of consensus"
repeat_region 58550. .58886
/note="L1MC4 repeat: matches 7261. .7565 of consensus"
repeat_region 58890. .59017
/note="L1MC4 repeat: matches 7699. .7827 of consensus"

Query Match 14.6%; Score 18; DB 9; Length 144536;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 agatgcctcacacctg 35
|||||
Db 141822 AGATGCTTACACACCTG 141805

RESULT 23
LOCUS AL359815 146571 bp DNA HTG 03-JUL-2001
DEFINITION Homo sapiens chromosome 1 clone RP11-63N8, *** SEQUENCING IN
ACCESSION AL359815
VERSION AL359815.11 GI:14596332
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 146571)
McLay, K.
Direct Submission
Submitted (02-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Jul 4, 2001 this sequence version replaced gi:14587003.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: ba63N8
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator ET-amersham; 6% of reads
Dye-terminator Big Dye; 93% of reads
Consensus quality: 143907 bases at least Q40
Consensus quality: 144683 bases at least Q30
Consensus quality: 145097 bases at least Q20
Insert size: 145671; sum-of-contigs
Insert size: 155726; 1.2% error; agarose-fp
Quality coverage: 7.35x in Q20 bases; sum-of-contigs
Quality coverage: 7.63x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
```

```

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 24804: contig of 24804 bp in length
24805 24904: gap of 100 bp
24905 35138: contig of 10234 bp in length
35139 35238: gap of 100 bp
35239 54888: contig of 19650 bp in length
54889 54988: gap of 100 bp
54989 69175: contig of 14187 bp in length
69176 69275: gap of 100 bp
69276 84233: contig of 14958 bp in length
84234 84333: gap of 100 bp
84334 99092: contig of 14759 bp in length
99093 99192: gap of 100 bp
99193 102346: contig of 3154 bp in length
102347 102446: gap of 100 bp
102447 116814: contig of 14368 bp in length
116815 116914: gap of 100 bp
116915 122988: contig of 6074 bp in length
122989 123088: gap of 100 bp
123089 146571: contig of 23483 bp in length.
Location/Qualifiers
1. 146571
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-63N8"
/clone_1fb="RP11-11.1"
1. 24804
/note="assembly_fragment:01506
fragment_chain:1
clone_end:T7
vector_side:left"
24905. .35138
/note="assembly_fragment:01749
fragment_chain:1"
35239. .54888
/note="assembly_fragment:01082
fragment_chain:1"
54989. .69175
/note="assembly_fragment:02954
fragment_chain:1"
69276. .84233
/note="assembly_fragment:01509"
84334. .99092
/note="assembly_fragment:02635
fragment_chain:2"
99193. 102346
/note="assembly_fragment:01308
fragment_chain:2"
102447. 116814
/note="assembly_fragment:02746
fragment_chain:2"
116915. 122988
/note="assembly_fragment:00077
fragment_chain:2"
123089. 146571
/note="assembly_fragment:01690
fragment_chain:2
clone_end:SP6
vector_side:right"

BASE COUNT 34096 a 36702 c 37423 g 37443 t 907 others

ORIGIN
Query Match 14.6%; Score 18; DB 2; Length 146571;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 aattatccacacaca 123
```

Db 85629 AATTATTCACACACCA 85646

RESULT 24

AC034225 148498 bp DNA HTG 19-APR-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2124P14, WORKING DRAFT
DEFINITION SEQUENCE, 8 ordered pieces.
ACCESSION AC034225.4 GI:9256727
VERSION AC034225.4
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 148498)
JOURNAL DOE Joint Genome Institute.
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 148498)
TITLE DOE Joint Genome Institute.
JOURNAL Direct Submission
COMMENT Submitted (05-APR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 18, 2000 this sequence version replaced gi:7712076.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 675539
Center clone name: CTRB-H1_2124P14

Summary Statistics
Consensus quality: 139048 bases at least Q40
Consensus quality: 145965 bases at least Q30
Consensus quality: 147145 bases at least Q20
Estimated insert size: 150000; pulse field gel estimation
Estimated insert size: 148148; sum-of-contigs estimation
Quality coverage: 5.89 in Q20 bases; pulse field gel estimation
Quality coverage: 5.89 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 8 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 61862: contig of 61862 bp in length
* 61863 61962: gap of unknown length
* 61963 68022: contig of 6060 bp in length
* 68023 68122: gap of unknown length
* 68123 80136: contig of 12014 bp in length
* 80137 80236: gap of unknown length
* 80237 83311: contig of 3075 bp in length
* 83312 83411: gap of unknown length
* 83412 88833: contig of 5422 bp in length
* 88834 88933: gap of unknown length
* 88934 92360: contig of 3427 bp in length
* 92361 92460: gap of unknown length
* 92461 120402: contig of 27942 bp in length
* 120403 120502: gap of unknown length
* 120503 148498: contig of 27996 bp in length.
Location/Qualifiers
1. 148498
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="CTD-2124P14"

BASE COUNT 44488 a 30209 c 29485 g 43610 t 706 others
ORIGIN

Query Match 14.6%; Score 18; DB 2; Length 148498;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 28 caaccctgacacctaat 45
|||||
Db 3264 CAAOCCTGAGACCTTAT 32661

RESULT 25

AL357252 153405 bp DNA HTG 02-MAY-2001
LOCUS Homo sapiens chromosome 1 clone RP11-35C9, *** SEQUENCING IN
DEFINITION PROGRESS ***; 10 unordered pieces.
ACCESSION AL357252
VERSION AL357252.8 GI:13990058
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 153405)
JOURNAL Direct Submission
COMMENT Submitted (01-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
On May 7, 2001 this sequence version replaced gi:12227371.
-----Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquerry@sanger.ac.uk
-----Project Information
Center project name: BA35C9

COMMENT

Assembly program: XMAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 150971 bases at least Q40
Consensus quality: 151464 bases at least Q30
Consensus quality: 151889 bases at least Q20
Insert size: 152505; sum-of-contigs
Insert size: 157709; 4.7% error; agarose-fp
Quality coverage: 7.06x in Q20 bases; sum-of-contigs Quality
coverage: 6.89x in Q20 bases; agarose-fp
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 21851: contig of 21851 bp in length
* 21852 21951: gap of 100 bp
* 21952 36986: contig of 15035 bp in length
* 36987 37086: gap of 100 bp
* 37087 51405: contig of 14319 bp in length
* 51406 51505: gap of 100 bp
* 51506 54730: contig of 3225 bp in length
* 54731 54830: gap of 100 bp
* 54831 75701: contig of 20871 bp in length
* 75702 75801: gap of 100 bp
* 75802 85440: contig of 9639 bp in length
* 85441 85540: gap of 100 bp
* 85541 116412: contig of 30872 bp in length

```

* 116413 116512: gap of 100 bp
* 116513 141040: contig of 24528 bp in length
* 141041 141140: gap of 100 bp
* 141141 145417: contig of 4277 bp in length
* 145418 145517: gap of 100 bp
* 145518 153405: contig of 7888 bp in length.
FEATURES
    source
        1. 153405
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="1"
            /clone="RP11-35C9"
            /clone_1fb="RP11-11.1"
            1. 21851
                /note="assembly_fragment:00790
                fragment_chain:1
                clone_end:17
                vector_side:left"
            21952. 36986
                /note="assembly_fragment:02366
                fragment_chain:1"
            37087. 51405
                /note="assembly_fragment:02196
                fragment_chain:1"
            51506. 54730
                /note="assembly_fragment:02549
                fragment_chain:1"
            54831. 75701
                /note="assembly_fragment:02618
                fragment_chain:1"
            75802. 85440
                /note="assembly_fragment:01594
                fragment_chain:1"
            85541. 116412
                /note="assembly_fragment:00526
                fragment_chain:2"
            116513. 141040
                /note="assembly_fragment:02336
                fragment_chain:2"
            141141. 145417
                /note="assembly_fragment:02107
                fragment_chain:2"
            145518. 153405
                /note="assembly_fragment:02012
                fragment_chain:2"
BASE COUNT  33074 a 40678 c 42451 g 36301 t 901 others
ORIGIN
Query Match      14.6%; Score 18; DB 2; Length 153405;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 106 aattattccacacaca 123
|||||
Db 37931 AATTATTCACACACACA 37948

RESULT 26
AC023537 156795 bp DNA HTG 12-MAR-2000
LOCUS Homo sapiens chromosome 4 clone RP11-394N5 map 4, WORKING DRAFT
DEFINITION AC023537
SEQUENCE AC023537
ACCESSION AC023537.2 GI:7229813
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 156795)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 4, clone RP11-394N5

```

```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 156795)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Bouknighter,B., Brown,A., Burkett,G., Campolondo,A., Castle,A.,
Chopel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Dodge,S., Domono,M., Doyle,M.,
Fenster,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatord,A., Hottom,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McGhara,A., McKernan,K., McSheeters,R., Meldrum,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olyvar,T.M.,
Peterson,K., Pierre,N., Pisanl,C., Pollata,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome
RESEARCH Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 12, 2000 this sequence version replaced gi:6978216.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: I5543
Center clone name: 394.N.5
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 137441 bases at least Q40
Consensus quality: 147789 bases at least Q30
Consensus quality: 151521 bases at least Q20
Insert size: 167000; agarose-fp
Insert size: 154095; sum-of-contigs
Quality coverage: 3.2 in Q20 bases; sum-of-contigs
Quality coverage: 3.5 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 28 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1411: contig of 1411 bp in length
* 1412 1511: gap of 100 bp
* 1512 2957: contig of 1446 bp in length
* 2958 3057: gap of 100 bp
* 3058 5096: contig of 2039 bp in length
* 5097 5196: gap of 100 bp
* 5197 7059: contig of 1863 bp in length
* 7060 7159: gap of 100 bp
* 7160 9463: contig of 2304 bp in length
* 9464 9563: gap of 100 bp
* 9564 11760: contig of 2197 bp in length
* 11761 11860: gap of 100 bp
* 11861 13457: contig of 1597 bp in length
* 13458 13557: gap of 100 bp
* 13558 15559: contig of 2002 bp in length

```

```

* 15560 15659: gap of 100 bp
* 15660 18335: contig of 2676 bp in length
* 18336 18435: gap of 100 bp
* 18436 20122: contig of 1687 bp in length
* 20123 20222: gap of 100 bp
* 20223 22027: contig of 1805 bp in length
* 22028 22127: gap of 100 bp
* 22128 24224: contig of 2097 bp in length
* 24225 24324: gap of 100 bp
* 24325 26969: contig of 2645 bp in length
* 26970 27069: gap of 100 bp
* 27070 30436: contig of 3367 bp in length
* 30437 30536: gap of 100 bp
* 30537 34800: contig of 4264 bp in length
* 34801 34900: gap of 100 bp
* 34901 38851: contig of 3951 bp in length
* 38852 38951: gap of 100 bp
* 38952 44086: contig of 5135 bp in length
* 44087 44186: gap of 100 bp
* 44187 449895: contig of 5709 bp in length
* 44986 49955: gap of 100 bp
* 49956 57232: contig of 7237 bp in length
* 57233 57332: gap of 100 bp
* 57333 64959: contig of 7627 bp in length
* 64960 65059: gap of 100 bp
* 65060 72868: contig of 7809 bp in length
* 72869 72968: gap of 100 bp
* 72969 81145: contig of 8177 bp in length
* 81146 81245: gap of 100 bp
* 81246 88655: contig of 7410 bp in length
* 88656 88755: gap of 100 bp
* 88756 97086: contig of 8331 bp in length
* 97087 97186: gap of 100 bp
* 97187 106956: contig of 9770 bp in length
* 106957 107056: gap of 100 bp
* 107057 119495: contig of 12439 bp in length
* 119496 119595: gap of 100 bp
* 119596 134366: contig of 14771 bp in length
* 134367 134466: gap of 100 bp
* 134467 156795: contig of 22329 bp in length.

```

FEATURES

```

source
1. 156795
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone-lib="RP11-394N5"
1. 1411
/note="assembly-fragment"
misc_feature
1512. 2957
/note="assembly-fragment"
misc_feature
3058. 5096
/note="assembly-fragment"
misc_feature
5197. 7059
/note="assembly-fragment"
misc_feature
7160. 9463
/note="assembly-fragment"
misc_feature
9564. 11760
/note="assembly-fragment"
misc_feature
11861. 13457
/note="assembly-fragment"
misc_feature
13558. 15559
/note="assembly-fragment"
misc_feature
15660. 18335
/note="assembly-fragment"
misc_feature
18436. 20122
/note="assembly-fragment"
misc_feature
20223. 22027
/note="assembly-fragment"
misc_feature
22128. 24224
/note="assembly-fragment"
24325. 26969

```

```

misc_feature /note="assembly-fragment"
27070. 30436 /note="assembly-fragment"
misc_feature 30537. 34800 /note="assembly-fragment"
misc_feature 34901. 38851 /note="assembly-fragment"
misc_feature clone_end:77
vector_side:left"
misc_feature 38952. 44086 /note="assembly-fragment"
misc_feature 44187. 449895 /note="assembly-fragment"
misc_feature 49956. 57232 /note="assembly-fragment"
misc_feature 57333. 64959 /note="assembly-fragment"
misc_feature 65060. 72868 /note="assembly-fragment"
misc_feature 72969. 81145 /note="assembly-fragment"
misc_feature 81246. 88655 /note="assembly-fragment"
misc_feature 88756. 97086 /note="assembly-fragment"
misc_feature 97187. 106956 /note="assembly-fragment"
misc_feature 107057. 119495 /note="assembly-fragment"
misc_feature 119596. 134366 /note="assembly-fragment"
clone_end:SP6
vector_side:left"
misc_feature 134467. 156795 /note="assembly-fragment"
BASE COUNT 47744 a 27616 c 27899 g 50833 t 2703 others
ORIGIN

```

```

Query Match 14.6%: Score 18; DB 2; Length 156795;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 28 caacctgatgacctaat 45
|||||
Db 146844 CAACCTGATGACCTAAT 146827

```

```

RESULT 27
AC024315/c DNA HTG 26-MAY-2000
LOCUS AC024315.3 165858 bp
DEFINITION Homo sapiens clone RP11-26N3, WORKING DRAFT SEQUENCE, 14 unordered
pieces.
ACCESSION AC024315
VERSION AC024315.3 GI:8072573
KEYWORDS HTG; PHASE1; HTGS; DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 165858)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 165858)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E.,
Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavsky,L.,
Bouckhaghter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
Fenesor,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,

```


TITLE
JOURNAL
COMMENT

Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Laroque, K., Lohocsky, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McSheeters, R., Meldrum, J., Menais, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wymann, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 25, 2000 this sequence version replaced g1:7239607.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www.seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L4668

Center clone name: 26_N_3

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 154942 bases at least Q40
Consensus quality: 159890 bases at least Q30
Consensus quality: 162505 bases at least Q20
Insert size: 164000; agarose-fp
Insert size: 164558; sum-of-ctrls
Quality coverage: 4.6 in Q20 bases; agarose-fp
Quality coverage: 4.6 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1      1274: contig of 1274 bp in length
*      1275 1374: gap of 100 bp
*      1375 2528: contig of 1154 bp in length
*      2529 2628: gap of 100 bp
*      2629 4506: contig of 1878 bp in length
*      4507 4606: gap of 100 bp
*      4607 6447: contig of 1841 bp in length
*      6448 6547: gap of 100 bp
*      6548 11790: contig of 5243 bp in length
*      11791 11890: gap of 100 bp
*      11891 18653: contig of 6763 bp in length
*      18654 18753: gap of 100 bp
*      18754 27335: contig of 8582 bp in length
*      27336 27435: gap of 100 bp
*      27436 41698: contig of 14263 bp in length
*      41699 41798: gap of 100 bp
*      41799 54767: contig of 12969 bp in length
*      54768 54867: gap of 100 bp
*      54868 70875: contig of 16008 bp in length
*      70876 70975: gap of 100 bp
*      70976 87982: contig of 17007 bp in length
*      87983 88082: gap of 100 bp
*      88083 110113: contig of 22031 bp in length
*      110114 110213: gap of 100 bp
*      110214 133760: contig of 23547 bp in length

```

```

*      133761 133860: gap of 100 bp
*      133861 165858: contig of 31998 bp in length.
FEATURES
Source
1..165858
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-26N3"
/clone_1lb="RPC1-11 Human Male BAC"

```

```

misc_feature
1..1274
/note="assembly_fragment"
misc_feature
1375..2528
/note="assembly_fragment"
misc_feature
2629..4506
/note="assembly_fragment"
misc_feature
4607..6447
/note="assembly_fragment"
misc_feature
6448..11790
/note="assembly_fragment"
misc_feature
11791..18653
/note="assembly_fragment"
misc_feature
18654..27335
/note="assembly_fragment"
misc_feature
27336..41698
/note="assembly_fragment"
misc_feature
41699..41798
/note="assembly_fragment"
misc_feature
41799..54767
/note="assembly_fragment"
misc_feature
54768..54867
/note="assembly_fragment"
misc_feature
54868..70875
/note="assembly_fragment"
misc_feature
70876..70975
/note="assembly_fragment"
misc_feature
70976..87982
/note="assembly_fragment"
misc_feature
87983..110113
/note="assembly_fragment"
misc_feature
110114..133760
/note="assembly_fragment"
misc_feature
133761..165858
/note="assembly_fragment"

```

```

BASE COUNT 48868 a 34483 c 34023 g 47184 t 1300 others
ORIGIN
Query Match 14.6% Score 18; DB 2; Length 165858;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 26 tacacacctgaccta 43
|||||
Db 116656 TACACCCCTGATGACCTA 116639

```

```

RESULT 28
AC009776
LOCUS Homo sapiens chromosome 12 clone RP11-333115, WORKING DRAFT
DEFINITION AC009776.8 GI:10445225
ACCESSION AC009776
VERSION AC009776
KEYWORDS HTG: HTGS_PHAS1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 183496)
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
Alstrooms, S.L., Amaralunge, H.C., Are, J.R., Banks, T., Barberia, J.,
Benton, J., Bimaye, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowle, S., Brieve, M., Brown, E., Brown, M., Bryant, N.P., Buhey, C.,
Burke, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F.,
Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chavez, G., Chen, R.,
Chen, Z., Chowdhury, I., Christopoulos, C., Cleveland, C.D., Cox, C.,

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Day-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Diaper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunartine, P., Hale, S.,
Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hayes, A.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Hollway, C.,
Hollins, B., Homs, J., Howard, S., Huber, J., Hulik, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Louisaged, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M.,
Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S.,
Ogulu, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Plickens, R., Primus, E., Pu, L.L.,
Rutiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoochari, N.,
Sisson, I., Sodergren, E., Sonalke, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Washington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 183496)
Worley, K.C.

Submitted (01-SEP-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Oct 2, 2000 this sequence version replaced gi:10047636.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMMH
Center clone name: RP11-333115
----- Summary Statistics
Sequencing vector: M13, L08821
Chemistry: Dye-terminator Bodipy: 61% of reads
Chemistry: Dye-terminator Big Dye: 39% of reads
Assembly program: Phrap: version 0.990139
Consensus quality: 165672 bases at least Q40
Consensus quality: 174319 bases at least Q30
Consensus quality: 178820 bases at least Q20
Estimated insert size: 176791: sum-of-ctrls estimation
Quality coverage: 3.8x in Q20 bases: sum-of-ctrls estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 29040: contig of 29040 bp in length
* 29041 29140: gap of unknown length
* 29141 48361: contig of 19221 bp in length

48362 48461: gap of unknown length
* 48462 65324: contig of 17063 bp in length
* 65325 65624: gap of unknown length
* 65625 82557: contig of 16933 bp in length
* 82558 82657: gap of unknown length
* 82658 99864: contig of 17207 bp in length
* 99865 99964: gap of unknown length
* 99965 112153: contig of 12189 bp in length
* 112154 112253: gap of unknown length
* 112254 123333: contig of 11080 bp in length
* 123334 123433: gap of unknown length
* 123434 133305: contig of 9872 bp in length
* 133306 133405: gap of unknown length
* 133406 139739: contig of 6334 bp in length
* 139740 139839: gap of unknown length
* 139840 147455: contig of 7616 bp in length
* 147456 147555: gap of unknown length
* 147556 152212: contig of 4657 bp in length
* 152213 152312: gap of unknown length
* 152313 157750: contig of 5438 bp in length
* 157751 157850: gap of unknown length
* 157851 161821: contig of 3971 bp in length
* 161822 161921: gap of unknown length
* 161922 165295: contig of 3374 bp in length
* 165296 165385: gap of unknown length
* 165386 165855: contig of 3190 bp in length
* 165856 168685: gap of unknown length
* 168686 171193: contig of 2508 bp in length
* 171194 171293: gap of unknown length
* 171294 173564: contig of 2271 bp in length
* 173565 173664: gap of unknown length
* 173665 176006: contig of 2342 bp in length
* 176007 176106: gap of unknown length
* 176107 177749: contig of 1643 bp in length
* 177750 177849: gap of unknown length
* 177850 180432: contig of 2583 bp in length
* 180433 180532: gap of unknown length
* 180533 182100: contig of 1568 bp in length
* 182101 182200: gap of unknown length
* 182201 183496: contig of 1296 bp in length.

----- Location/Qualifiers
source
1. 183496
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="12"
/clone="RP11-333115"

BASE COUNT 51162 a 39432 c 39724 g 51055 t 2123 others

ORIGIN

Query Match 14.6%; Score 18; DB 2; Length 183496;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 106 aattatctcacacaca 123
Db 180935 AATTATTCACACACA 180952

RESULT 29
AC068581
LOCUS AC068581 190663 bp DNA HTG 23-SEP-2000
DEFINITION Homo sapiens chromosome 4 clone RP11-596012 map 4, WORKING DRAFT
SEQUENCE, 22 unordered pieces.
ACCESSION AC068581
VERSION AC068581.2 GI:10280770
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 190663)
Britten, B., Linton, L., Nussbaum, C. and Lander, E.

TITLE Homo sapiens chromosome 4, clone RP11-598012
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 190663)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abramson, H., Allen, N., Anderson, S., Balwin, J., Barna, N., Bastien, V., Bedalov, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choehel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeArnell, K., Dewar, K., Diaz, J.S., Dodge, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardy, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heath, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Lacroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, G., Liu, G., Locke, K., Macdonald, P., Margulis, N., McCarthy, M., McKean, P., McGuck, A., McKernan, K., McPheters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivar, T.M., Oliver, J., Peterson, K., Pierre, N., Pisanil, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Testa, S., Theodore, J., Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE Direct Submission
JOURNAL Submitted (04-MAY-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Sep 23, 2000 this sequence version replaced g1:7705127. All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: LI0318

Center clone name: 598_O_12

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960711
Consensus quality: 177632 bases at least Q40
Consensus quality: 184698 bases at least Q20
Consensus quality: 186982 bases at least Q20
Insert size: 194000; agarose-fp
Insert size: 188563; sum-of-contigs
Quality coverage: 4.0 in Q20 bases; agarose-fp
Quality coverage: 4.1 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 22 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

* 1 4553: contig of 4553 bp in length
* 4554 4653: gap of 100 bp
* 4654 5713: contig of 1060 bp in length
* 5714 5813: gap of 100 bp
* 5814 9313: contig of 3500 bp in length
* 9314 9413: gap of 100 bp
* 9414 12245: contig of 2832 bp in length
* 12246 12345: gap of 100 bp
* 12346 15186: contig of 2841 bp in length
* 15187 15286: gap of 100 bp
* 15287 19162: contig of 3876 bp in length
* 19163 19262: gap of 100 bp
* 19263 23897: contig of 4635 bp in length
* 23898 23997: gap of 100 bp

```

FEATURES

source

```

* 23998 27485: contig of 3488 bp in length
* 27486 27585: gap of 100 bp
* 27586 31745: contig of 4160 bp in length
* 31746 31845: gap of 100 bp
* 31846 36461: contig of 4616 bp in length
* 36462 36561: gap of 100 bp
* 36562 42979: contig of 6418 bp in length
* 42980 43079: gap of 100 bp
* 43080 50663: contig of 7584 bp in length
* 50664 50763: gap of 100 bp
* 50764 58952: contig of 8189 bp in length
* 58953 59052: gap of 100 bp
* 59053 67775: contig of 8723 bp in length
* 67776 67875: gap of 100 bp
* 67876 77789: contig of 9914 bp in length
* 77790 77889: gap of 100 bp
* 77890 86920: contig of 9031 bp in length
* 86921 87020: gap of 100 bp
* 87021 110977: contig of 23957 bp in length
* 110978 111077: gap of 100 bp
* 11078 125542: contig of 14465 bp in length
* 125543 125642: gap of 100 bp
* 125643 141160: contig of 15518 bp in length
* 141161 141260: gap of 100 bp
* 141261 154817: contig of 13557 bp in length
* 154818 154917: gap of 100 bp
* 154918 176781: contig of 21864 bp in length
* 176782 176881: gap of 100 bp
* 176882 190663: contig of 13782 bp in length.

Location/Qualifiers
1. 190663
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-598012"
/clone_lib="RPC1-11 Human Male BAC"
1. 4553
/note="assembly-fragment"
clone_end:Sp6
vector_side:left"
4654. 5713
/note="assembly-fragment"
5814. 9313
/note="assembly-fragment"
9414. 12245
/note="assembly-fragment"
12346. 15186
/note="assembly-fragment"
15287. 19162
/note="assembly-fragment"
19263. 23897
/note="assembly-fragment"
23998. 27485
/note="assembly-fragment"
27586. 31745
/note="assembly-fragment"
31846. 36461
/note="assembly-fragment"
36562. 42979
/note="assembly-fragment"
43080. 50663
/note="assembly-fragment"
50764. 58952
/note="assembly-fragment"
59053. 67775
/note="assembly-fragment"
67876. 77789
/note="assembly-fragment"
77890. 86920
/note="assembly-fragment"
87021. 110977
/note="assembly-fragment"

```

```

misc_feature      111078..125542
                  /note="assembly-fragment"
misc_feature      125643..141160
                  /note="assembly-fragment"
misc_feature      141261..154817
                  /note="assembly-fragment"
misc_feature      154918..176781
                  /note="assembly-fragment"
misc_feature      176882..190663
                  /note="assembly-fragment"
                  clone_end:77
                  vector_side:right"
BASE COUNT      61633 a 34788 c 33742 g 58384 t 2116 others
ORIGIN

Query Match      14.6%; Score 18; DB 2; Length 190663;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      28 caaccctgatgacctaat 45
         |||||||
Db 164508 CAACCTGATGACTTAAT 164525

RESULT 30
AC079542      192430 bp      DNA      HTG      02-SEP-2000
DEFINITION    Mus musculus clone RP23-373B15, WORKING DRAFT SEQUENCE, 18
               unordered pieces.
ACCESSION     AC079542
VERSION       AC079542.1 GI:9964907
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE        house mouse.
ORGANISM      Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1 (bases 1 to 192430)
               DOE Joint Genome Institute.
               Sequencing of Mouse
               Unpublished
               2 (bases 1 to 192430)
               DOE Joint Genome Institute.
               Direct Submission
               Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
               Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
               -----Genome Center
               Center: Joint Genome Institute
               Center Code: JGI
               Web site: http://www.jgi.doe.gov
               -----
Project Information
Center Project Name: 1880143
Center clone name: RPCI-23_373B15
-----
Summary Statistics
Consensus quality: 177736 bases at least Q40
Consensus quality: 184740 bases at least Q30
Consensus quality: 186701 bases at least Q20
Estimated insert size: 192000; agarose-fp estimation
Estimated insert size: 190730; sum-of-contigs estimation
Quality coverage: 8.68 in Q20 bases; agarose-fp estimation
Quality coverage: 8.74 in Q20 bases; sum-of-contigs estimation
NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1123: contig of 1123 bp in length
* 1124 1223: gap of unknown length

```

```

* 1224 2294: contig of 1071 bp in length
* 2295 2394: gap of unknown length
* 2395 3567: contig of 1173 bp in length
* 3568 4937: gap of unknown length
* 4938 5037: gap of 1270 bp in length
* 5038 6776: gap of unknown length
* 6777 6876: contig of 1739 bp in length
* 6877 8755: gap of unknown length
* 8756 8855: contig of 1879 bp in length
* 8856 11215: gap of unknown length
* 11216 11315: contig of 2360 bp in length
* 11316 14746: gap of unknown length
* 14747 14846: contig of 3431 bp in length
* 14847 20982: gap of unknown length
* 20983 21082: contig of 6136 bp in length
* 21083 27586: gap of unknown length
* 27587 27686: contig of 6504 bp in length
* 27687 36183: gap of unknown length
* 36184 36283: contig of 8437 bp in length
* 36284 45893: gap of unknown length
* 45894 45993: contig of 9610 bp in length
* 45994 58318: gap of unknown length
* 58319 58418: contig of 12325 bp in length
* 58419 79391: gap of unknown length
* 79392 79491: contig of 20973 bp in length
* 79492 105308: gap of unknown length
* 105309 105408: contig of 25817 bp in length
* 105409 124964: gap of unknown length
* 124965 125064: gap of 19556 bp in length
* 125065 152452: gap of unknown length
* 152453 152552: gap of 27388 bp in length
* 152553 192430: contig of unknown length
* 192430 39878 bp in length.
Location/Qualifiers
source
  /organism="Mus musculus"
  /db_xref="taxon:10090"
  /clone="RP23-373B15"
  /clone_lib="RPCI mouse BAC library 23"
BASE COUNT      48967 a 46656 c 46744 g 48326 t 1737 others
ORIGIN

Query Match      14.6%; Score 18; DB 2; Length 192430;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      56 gagggctgactgactgg 73
         |||||||
Db 134515 GAGGGCTGCACCTGCTGG 134532

RESULT 31
AL357568      198605 bp      DNA      HTG      03-MAY-2001
DEFINITION    Homo sapiens chromosome 1 clone RP11-506O24, *** SEQUENCING IN
               PROGRESS ***, 8 unordered pieces.
ACCESSION     AL357568
VERSION       AL357568.11 GI:13990061
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE        human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 198605)
               Burton, J.
               Direct Submission
               Submitted (03-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
               CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
               Requests: clonequests@sanger.ac.uk
               On May 7, 2001 this sequence version replaced gi:12329367.
               -----
               Genome Center
               Center: Sanger Centre

```

Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquerry@sanger.ac.uk
 ----- Project Information
 Center project name: BA506024
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; 108752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 197163 bases at least Q40
 Consensus quality: 197163 bases at least Q30
 Consensus quality: 197477 bases at least Q20
 Insert size: 197905; sum-of-contrigs
 Insert size: 138913; 37.8% error; agarose-fp
 Quality coverage: 7.56x in Q20 bases; sum-of-contrigs Quality
 coverage: 10.97x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contrigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contrigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1 25045: contrig of 25045 bp in length
 * 25046 25145: gap of 100 bp
 * 25146 33473: contrig of 8328 bp in length
 * 33474 33573: gap of 100 bp
 * 33574 79364: contrig of 45791 bp in length
 * 79365 79464: gap of 100 bp
 * 79465 104064: contrig of 24600 bp in length
 * 104065 104164: gap of 100 bp
 * 104165 129888: contrig of 25724 bp in length
 * 129889 129988: gap of 100 bp
 * 129989 139533: contrig of 9545 bp in length
 * 139534 139633: gap of 100 bp
 * 139634 157494: contrig of 17861 bp in length
 * 157495 157594: gap of 100 bp
 * 157595 198605: contrig of 41011 bp in length.
 Location/Qualifiers
 1. 198605
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-506024"
 /clone_11b="RP11-11.2"
 1. 25045
 /note="assembly-fragment:02045
 fragment.chain:1"
 misc-feature
 25146 33473
 /note="assembly-fragment:03085
 fragment.chain:1"
 misc-feature
 33574 79364
 /note="assembly-fragment:02873
 fragment.chain:1"
 misc-feature
 79465 104064
 /note="assembly-fragment:01381
 fragment.chain:1"
 misc-feature
 104165 129888
 /note="assembly-fragment:00526.0"
 misc-feature
 129989 139533
 /note="assembly-fragment:00526.1"
 misc-feature
 139634 157494
 /note="assembly-fragment:02854.0"
 misc-feature
 157595 198605
 /note="assembly-fragment:03653"
 BASE COUNT 57345 a 40525 c 41243 g 58791 t 701 others
 ORIGIN

Query Match 14.6%; Score 18; DB 2; Length 198605;
 Best Local Similarity 100.0%; Pred. No. 15;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 26 tacaccctatgaccta 43
 ||||||||||||||||
 Db 96580 TACACCCTATGACCTA 96597

RESULT 32
 AC025501
 LOCUS
 DEFINITION
 AC025501
 AC025501
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 218073)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Mus musculus chromosome 10, clone RP23-129023
 Unpublished
 2 (bases 1 to 218073)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
 Boguslavsky,L., Bouknight,B., Brown,A., Burkett,G.,
 Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
 Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
 Grand-Pierre,N., Grant,G., Hages,B., Heatford,A., Horton,L.,
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
 Klein,J., Lakoque,K., Lamazares,R., Lander,T., Lehoczy,J.,
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPeeters,R.,
 Meldrum,J., Meneses,L., Mihova,T., Miranda,C., Miene,V., Morrow,J.,
 Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Olyar,T.M., Oliver,J., Peterson,K., Pierre,N.,
 Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
 Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
 Testafye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
 Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
 Young,G., Zainoun,J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (09-MAR-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 25, 2001 this sequence version replaced gi:8077096.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RV/RepeatMasker.html>

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L8063
 Center clone name: 129_O_23
 ----- Summary Statistics
 Sequencing vector: M13; M7815; 44% of reads
 Sequencing vector: Plasmid; n/a; 56% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 213014 bases at least Q40
 Consensus quality: 214764 bases at least Q30
 Consensus quality: 215528 bases at least Q20
 Insert size: 210000; agarose-fp
 Insert size: 216373; sum-of-contrigs
 Quality coverage: 10.2 in Q20 bases; agarose-fp
 Quality coverage: 9.9 in Q20.
 * NOTE: This is a 'working draft' sequence. It currently

```

* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 23321: contig of 23321 bp in length
23322 23421: gap of 100 bp
23422 24529: contig of 1108 bp in length
24530 24629: gap of 100 bp
24630 26959: contig of 2330 bp in length
26960 27059: gap of 100 bp
27060 28240: contig of 1181 bp in length
28241 28340: gap of 100 bp
28341 30420: contig of 2080 bp in length
30421 30520: gap of 100 bp
30521 32131: contig of 1611 bp in length
32132 32231: gap of 100 bp
32232 34170: contig of 1939 bp in length
34171 34270: gap of 100 bp
34271 36716: contig of 2446 bp in length
36717 36816: gap of 100 bp
36817 40349: contig of 3533 bp in length
40350 40449: gap of 100 bp
40450 43365: contig of 2916 bp in length
43366 43465: gap of 100 bp
43466 46874: contig of 3409 bp in length
46875 46974: gap of 100 bp
46975 51691: contig of 4717 bp in length
51692 51791: gap of 100 bp
51792 118666: contig of 66875 bp in length
118667 118766: gap of 100 bp
118767 133750: contig of 14984 bp in length
133751 133850: gap of 100 bp
133851 151236: contig of 17386 bp in length
151237 151336: gap of 100 bp
151337 174475: contig of 23139 bp in length
174476 174575: gap of 100 bp
174576 208697: contig of 34122 bp in length
208698 208797: gap of 100 bp
208798 218073: contig of 9276 bp in length.
Location/Qualifiers
1. 218073
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="10"
/map="10"
/clone="RP23-129023"
/clone_lib="RPCI-23 Female Mouse BAC"
1. 23321
/note="assembly-fragment"
clone_end:sp6
vector_side:left"
misc_feature
24630..26959
/note="assembly-fragment"
misc_feature
27060..28240
/note="assembly-fragment"
misc_feature
28341..30420
/note="assembly-fragment"
misc_feature
30521..32131
/note="assembly-fragment"
misc_feature
32232..34170
/note="assembly-fragment"
misc_feature
34271..36716
/note="assembly-fragment"
misc_feature
36817..40349
/note="assembly-fragment"
misc_feature
40450..43365
/note="assembly-fragment"
misc_feature
43466..46874
/note="assembly-fragment"

```

```

misc_feature /note="assembly-fragment"
46875..51691
misc_feature /note="assembly-fragment"
51792..118666
misc_feature /note="assembly-fragment"
118667..133750
misc_feature /note="assembly-fragment"
133851..151236
misc_feature /note="assembly-fragment"
151337..174475
misc_feature /note="assembly-fragment"
174576..208697
misc_feature /note="assembly-fragment"
208798..218073
vector_side:right"
clone_end:T7
BASE COUNT 61957 a 47778 c 47469 g 59125 t 1744 others
ORIGIN
Query Match 14.6%; Score 18; DB 2; Length 218073;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 74 taagttcacatgtaag 91
Db 133435 TAACTTCACATGTAAAG 133452
RESULT 33
AC079583/c
LOCUS AC079583
DEFINITION Mus musculus clone RP23-9G1, WORKING DRAFT SEQUENCE, 19 unordered
pieces.
ACCESSION AC079583
VERSION AC079583.1 GI:964948
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 226060)
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 226060)
DOE Joint Genome Institute.
Direct Submission
Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1740473
Center clone name: RPCI-23_9G1
-----
Summary Statistics
Consensus quality: 209212 bases at least Q40
Consensus quality: 215597 bases at least Q30
Consensus quality: 216943 bases at least Q20
Estimated insert size: 212000; agarose-fp estimation
Estimated insert size: 224260; sum-of-contigs estimation
Quality coverage: 10.49 in Q20 bases; agarose-fp estimation
Quality coverage: 9.92 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```



```

DE Heterocapsa pygmaea strain CCMP 1490 psba gene, complete cds; chloroplast
DE gene for chloroplast product.
XX
XX
XX Heterocapsa pygmaea
XX Eukaryota; Alveolata; Dinophyceae; Peridinales; Heterocapsaceae;
XX Heterocapsa.
XX Plastid:Chloroplast
XX
XX {1}
XX 1-2195
XX Zhang Z., Green B.R., Cavalier-Smith T.;
XX "Evolution and origin of dinoflagellate chloroplast mitochondria
XX chromosomes";
XX Unpublished.
XX
XX {2}
XX 1-2195
XX Zhang Z., Green B.R., Cavalier-Smith T.;
XX
XX Submitted (29-JUN-2000) to the EMBL/GenBank/DBJ databases.
XX Department of Botany, The University of British Columbia, #3529-6270
XX University Blvd., Vancouver, BC V6T 1Z4, Canada
XX
XX SPTREMBL; Q9MSC1; Q9MSC1.
XX
XX Key Location/Qualifiers
XX
XX source 1..2195
XX /db_xref="taxon:35672"
XX /organism="plastid:chloroplast"
XX /organism="Heterocapsa pygmaea"
XX /strain="CCMP 1490"
XX 826..1872
XX /codon_start=1
XX /db_xref="SPTREMBL:Q9MSC1"
XX /product="psba"
XX /protein_id="AAG25876.1"
XX /translation="MKNTFNTSNFANAYSEWYIGFTLSTNSRLYIGMFGILMEPL
XX VLATVAYIAAFIAPPVIDGIREPVAGALLYNNISGAVIPSSNAIGVHYVWEAL
XX GFDEMLYNGGYOFVLFHRTIGAGAYGREMEFARLGMRFIVAFSAPIVAASAVFI
XX VYPIGGSFSDGMPICIGSTFPMFLVROAENHILMHPHILICVAAYFGSLSANHGS
XX VTSSLVATACDLSLVNGYNGOEDETISANAGTFGRLLIQVAFNSRSLSHFFLAA
XX WPIVIGWETALGVSTMAFNLNGLNFGNOSIIDSGLINSWADIWNRADLGMEVHERNA
XX HNFPLDLA"
XX
XX Sequence 2195 BP; 591 A; 383 C; 445 G; 776 T; 0 other;
XX
XX
XX Query Match 13.8%; Score 17; DB 20; Length 2195;
XX Best Local Similarity 100.0%; Pred. No. 93;
XX Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 101 tctcaatttccac 117
XX ||||||||||||||||
XX Db 565 TCTCAATTATCCAC 549
XX
XX
XX RESULT 37
XX CELYS8G8A 34544 bp DNA INV 28-MAR-2000
XX LOCATION Caenorhabditis elegans cosmid Y58G8A, complete sequence.
XX ACCESSION AC006808
XX VERSION AC006808.1 GI:4263130
XX KEYWORDS HTG.
XX SOURCE Caenorhabditis elegans.
XX ORGANISM Caenorhabditis elegans.
XX Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
XX Rhabditoidea; Rhabditidae; Pelodidae; Caenorhabditis.
XX
XX REFERENCE
XX 1 (bases 1 to 34544)
XX The C. elegans Genome Sequencing Consortium, Washington University
XX Genome Sequencing Center, St. Louis U.S.A. and the Sanger Centre,

```

```

TITLE Hinxton, U.K.,C.
JOURNAL Genome sequence of the nematode C. elegans: a platform for
MEDLINE Investigating Biology. The C. elegans Sequencing Consortium
99069613 Science 282 (5396), 2012-2018 (1998)
REMARK Erratum:[published errata appear in Science 1999 Jan
1:283(5398):35 and 1999 Mar 26:283(5410):2103 and 1999 Sep
3:285(5433):1493]]
REFERENCE
AUTHORS 2 (bases 1 to 34544)
TITLE Tin-Mollam,A., Graves,T. and Harrison,M.
JOURNAL The sequence of C. elegans cosmid Y58G8A
AUTHORS Unpublished
TITLE 3 (bases 1 to 34544)
JOURNAL Waterston,R.H.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (23-FEB-1999) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS 4 (bases 1 to 34544)
TITLE Waterston,R.H.
JOURNAL Direct Submission
TITLE Submitted (01-MAR-2000) Genome Sequencing Center, Washington
JOURNAL University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS 5 (bases 1 to 34544)
TITLE Waterston,R.
JOURNAL Direct Submission
TITLE Submitted (28-MAR-2000) Department of Genetics, Washington
JOURNAL University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA, and
Sanger Centre, Hinxton Hall
Cambridge CB10 1RO, England
e-mail: twinematode.wustl.edu and jessesanger.ac.uk

```

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

NEIGHBORING COSMID INFORMATION:

This is a segment of the YAC Y58G8, sequenced to span the gap between R02C2 and F33E11. The 5' clone is R02C2, 3100 bp overlap; 3' clone is F33E11, 200 bp overlap. Actual start of this YAC is at base position 20847 of CELDC2; actual end is at 7365 of CELT22H9. Sequence fidelity from base pair 6911 to base pair 9100 can not be guaranteed due to a tandem repeat. This region may contain misassemblies, single cloned areas, or regions of low quality.

NOTES:

Coding sequences below are predicted from computer analysis, using the program GeneFINDER (P. Green and L. Hillier, ms in preparation).

```

FEATURES
source
1..34544
/organism="Caenorhabditis elegans"
/strain="Bristol N2"
/db_xref="taxon:6239"
/chromosome="V"

```



```

gene      /clone="Y58G8A"
          /gene="Y58G8A.2"
          join(2379. .2443,2939. .3603,3779. .5358)
CDS       /gene="Y58G8A.2"
          /note="contains similarity to reverse transcriptases"
          /codon_start=1
          /evidence=not_experimental
          /protein_id="AAF60814.1"
          /db_xref="GI:7332127"
          /translation="MRKENKLCITRGDRNFEEYQSKYKLIIGDENARVGNKRDENE
          RYGTNAMEPRNDTGEIATFCESENRLMTNSQFYPIRRRTWVSPDGMKEIDINI
          LANGKFTVDTSVLPSEPTNGSDHRLSRNHNFNKLKLOVRRKRPKRLKLEATLA
          VSTTIAAPHGIDDEDYKIRVDSIKAAOSQSVYQPSNANRNLPELRRRLM
          RSDPNKSLSTICREAVERRDHENAKVEKHLPOLPPPTPIAEPLPREIREIAA
          INSFPNGKAAGSDKITTADFLKSGDNVILITGRNRTLESGETPDMWTSKTLTFK
          KQDRELENYRPICLPLVLYKTFKCVLRIRRTLEADPEVQGRSRSPSTIDHIS
          VQRLLENGREYQIPLTLVEIDQKAFDSVEPODWDLSQOQVQNGYINLQECYTD
          STFTFPHKKNVPIVIRGROGPISPNLFACIEHVFQLMKHKRGDERYTEGIR
          VNGOQLTNREFADIVLVANHPRTASOMTELEKSSVGLKINTGKTVLRNRTFQ
          RKEIRCPMTNTIIDDVNEYIYLGROIINSNLLPELHRRRAAANAFTNKRSTDO
          ITCPRLRALPDSTVLPATYGESEANTFKELAERYVTHALERLVLTLTEQER
          NTHREVRKSKLRDPLHITKKKLGAGHVARIDGRMTTLMNDWCPDERKPVGR
          PMRNDSLKKEVYTRDAFGQVQLMSTQA"
          10191. .13116
gene      /gene="Y58G8A.1"
          join(10191. .10218,10271. .10440,10953. .11104,11151. .11292,
          11346. .11490,11940. .12106,12166. .12319,12367. .12556,
          12804. .13116)
          /gene="Y58G8A.1"
          /note="contains similarity to Pfam family PF00065
          (neurotransmitter-gated ion-channel), Score=37.5,
          E=3.2e-10, N=2; coded for by C. elegans cDNA yk304h6.3;
          coded for by C. elegans cDNA yk304h6.5; coded for by C.
          elegans cDNA yk468h10.3; coded for by C. elegans cDNA
          yk468h10.5"
          /codon_start=1
          /protein_id="AAF60813.1"
          /db_xref="GI:7332126"
          /translation="MILRLISIGVILNHFQNAASQSDERKLEQLKQVSKYRP
          VKVESTVQVAVYLNIAHVEKDEHQTALVHGHMASWTDLELKERATNISTISI
          ASSKLWQPALQLYNSARGNSWSLYLGLPAIYVNGKWSRGTFSPFCQDPETWP
          YDQSCPIVITDWWYGLQVNSLSDPATAGYGPRTIRLSYDDETFNKRKHVGMEVD
          SMKHCYMGPKGCLDAEPDGNPDMWSYLEFGITLKRHLPYFSLTIVMPVSTLML
          LCFMSENFELRVFLITAIATAGVSNLISMPGSGRTPKRVHIXSLGISASQV
          VLIYMNELRVLPSPHYERDEGCDITRIPKKSGLTFPTTKGISEPQELNQAQTS
          EALIQEASTSIGISNPLAESGSDTDLNLNPNSEADISERSETLGMPSSEKTOES
          QKMRKLEIQLAIKRAFFEVLYFYLAFLVLAAT"
          complement(20844. .22820)
          /gene="Y58G8A.3"
          complement(join(20844. .20893,20937. .21123,21220. .21426,
          22088. .22244,22429. .22620,22807. .22820))
          /gene="Y58G8A.3"
          /note="contains similarity of Pfam family PF00001 (7
          transmembrane receptor (rhodopsin)), Score=132.4,
          E=5.1e-41, N=1"
          /codon_start=1
          /evidence=not_experimental
          /protein_id="AAF60815.1"
          /db_xref="GI:7332128"
          /translation="MSAALDFYIRSIFFLYGLFVYLGIFGNGVLAVANRKLQSA
          RNVFLNLIFDTLIEVTAIPYPMATKDMAFSGVMCHVILPMSCEVYASMLT
          AISDKFLHINDPTKQPSIRQALATPFLIMVSTLILNIPYLSPEHNGSEFVQGE
          TPYCGHFCDEANMOSNSRKITYGTITMLQFVPMVAVITTYCTFKILQKSKDITONA
          QPCQSLTQKQSDATSRKKRVNILLIIMAVVTFLGISLYLFFGKNSQLF"
          complement(23241. .24074)
          /gene="Y58G8A.4"
          complement(join(23241. .23565,23665. .23873,23982. .24074))
          /gene="Y58G8A.4"
          /codon_start=1
          /evidence=not_experimental
          /product="Hypothetical protein Y58G8A.4"
          /protein_id="AAF60816.1"

```

```

          /db_xref="GI:7332129"
          /translation="MELIICFLWVLAHFKAGPEIDLNFENITNLYKKPFEPRLKED
          RKIDPMINONTLEPDIIEKNIIFRSNLKNLTKTFSTFLEFYGYACFQWNVIT
          FDMVLVRPPEPQORLPPEPSATLSQVITRWAFRSIRLRHRSASMLLCKFWVS
          WMDFLYVNOKNSTFLTKITISVIVORSHLPHRPLERLVKHPNF"
BASE COUNT 12193 a 5567 c 6036 g 10748 t
ORIGIN
Query Match 13.8%; Score 17; DB 3; Length 34544;
Best Local Similarity 100.0%; Pred. No. 66;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 101 |tttcaattattttccc|
Db 34189 |tcttcaattattttccac| 34205

RESULT 38
AL591603/c 38225 bp DNA PRI 18-JUL-2001
LOCUS AL591603
DEFINITION Human DNA sequence from clone RP11-320P15 on chromosome 6, complete
sequence.
ACCESSION AL591603
VERSION AL591603.4 GI:14970814
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 38225)
REFERENCE
AUTHORS Clark, G.
TITLE Direct Submission
JOURNAL Submitted (18-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
COMMENT
On Jul 19, 2001 this sequence version replaced gi:14329623.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Emr, EMBL; Sw,
SWISSPROT; Tr, TrEMBL; Wp, WormPeP; information on the WormPeP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
RP11-320P15 is from the library RPc1-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-320P15. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone RP11-100E6 is at 36226 in this sequence.
The true right end of clone RP11-239I6 is at 2000 in this sequence.
FEATURES
source
1. .36225
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"

```

```

/clone="RP11-320P15"
/clone_lib="RP11-11.2"
repeat_region
/note="LIM4 repeat: matches 4735. .4952 of consensus"
266. .368
/note="LIMC1 repeat: matches 6225. .6318 of consensus"
359. .563
/note="LIM1A1 repeat: matches 1. .218 of consensus"
564. .918
/note="RHE1B repeat: matches 1. .364 of consensus"
919. .1073
/note="LIM1A1 repeat: matches 218. .365 of consensus"
1076. .1388
/note="LIMC1 repeat: matches 5883. .6227 of consensus"
1390. .1733
/note="LIM1A1 repeat: matches 1. .426 of consensus"
1737. .2398
/note="LIMC1 repeat: matches 5219. .5883 of consensus"
2372. .2594
/note="LIM4 repeat: matches 4490. .4731 of consensus"
2628. .4554
/note="LIM4 repeat: matches 2176. .4208 of consensus"
4590. .9234
/note="LIMB2 repeat: matches 1432. .6115 of consensus"
9330. .9451
/note="61 copies 2 mer tt 59% conserved"
9690. .11001
/note="LIMC repeat: matches 653. .2052 of consensus"
11318. .11442
/note="LIM repeat: matches 63. .184 of consensus"
11802. .13026
/note="L2 repeat: matches 1035. .2358 of consensus"
13769. .14132
/note="RHE1C repeat: matches 3. .359 of consensus"
16645. .16967
/note="LIR1C repeat: matches 50. .383 of consensus"
18546. .18902
/note="L2 repeat: matches 2365. .2710 of consensus"
19303. .19360
/note="MER91 repeat: matches 1. .63 of consensus"
19987. .20040
/note="MER20 repeat: matches 160. .214 of consensus"
20092. .20224
/note="MIR repeat: matches 95. .221 of consensus"
20332. .20635
/note="ALUSP repeat: matches 1. .305 of consensus"
20787. .20882
/note="L2 repeat: matches 2665. .2700 of consensus"
21017. .21469
/note="MER89 repeat: matches 8. .557 of consensus"
21997. .22376
/note="MER57B repeat: matches 42. .403 of consensus"
25114. .25468
/note="LIM1A2 repeat: matches 1. .374 of consensus"
26081. .26391
/note="ALU repeat: matches 1. .311 of consensus"
26905. .27023
/note="MIR repeat: matches 46. .175 of consensus"
28243. .28715
/note="LIM4 repeat: matches 5817. .6300 of consensus"
29703. .29873
/note="MLT1H repeat: matches 91. .271 of consensus"
29991. .30151
/note="MIR repeat: matches 28. .186 of consensus"
30434. .30744
/note="ALU repeat: matches 2. .311 of consensus"
31502. .31823
/note="L1 repeat: matches 4734. .5064 of consensus"
32200. .32235
/note="LIMC/D repeat: matches 5528. .5563 of consensus"
32330. .32461
/note="L1 repeat: matches 3821. .3954 of consensus"
32465. .32859
repeat_region

```

```

repeat_region /note="LIR33 repeat: matches 101. .520 of consensus"
32933. .33053
/note="L1 repeat: matches 3692. .3816 of consensus"
repeat_region 33957. .34033
/note="LIMC/D repeat: matches 5464. .5553 of consensus"
34424. .34729
/note="LIM4 repeat: matches 4696. .5015 of consensus"
35810. .35863
/note="27 copies 2 mer ac 70% conserved"
BASE COUNT 11082 a 6661 c 6907 g 13575 t
ORIGIN
Query Match 13.8%; Score 17; DB 9; Length 38225;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 29 aaccctgatacctaata 45
Db 1498 AACCTGATGACCTAAT 1482
|||||

```

```

RESULT 39
SPEC16C4/c 44751 bp DNA PLN 10-JUN-2000
LOCUS S.pombe chromosome III cosmid c16c4.
DEFINITION AL031535
ACCESSION AL031535.1 GI:3560254
KEYWORDS 60s ribosomal protein L12A; cdc2-cdc8 family serine-threonine
protein kinase; G-beta repeats; gene free region;
glucosamine-6-phosphate isomerase; low complexity gene free region;
n-terminal acetyltransferase; peptidyl-prolyl cis-trans isomerase;
pseudouridylylate synthase; RNA binding protein; rna binding protein;
rp12.1; sol1 family protein; st5; target of the inhibitor
staurosporine; ttf2 type LRR; TPR Domain; transcription factor
subunit; transcription initiation factor ttfid 60 kd subunit; WD
repeat domain.
fission yeast.

```

SOURCE Schizosaccharomyces pombe
ORGANISM Schizosaccharomyces pombe
Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
Schizosaccharomycetales; Schizosaccharomycetaceae;

REFERENCE 1 (bases 1 to 44751)
AUTHORS Purnelle,B., Goffeau,A., Wood,V., Rajandream,M.A. and Barrell,B.G.
TITLE Direct Submission
JOURNAL Submitted (03-SEP-1998) European Schizosaccharomyces genome
sequencing project, Sanger Centre, The Wellcome Trust Genome
Campus, Hinxton, Cambridge CB10 1SA, E-mail: Barrell@sanger.ac.uk
and Unite de Biochimie Physiologique, Universite Catholique de
Louvain, Place Croix du Sud 2/20, B-1348 Louvain-la-Neuve, Belgium
Notes:

COMMENT

Details of yeast sequencing at the Sanger Centre are available on the World Wide Web.
(URL, <http://www.sanger.ac.uk/projects/S-pombe/>)
During 1995 to 1996 about 66% of S. pombe chromosome 1 was sequenced by the Sanger Centre. The sequencing of the S. pombe genome is now being continued with funding from the European Commission. Fourteen European sequencing laboratories, including the Sanger Centre, are participating in the project.
Protein coding regions (CDS) have been predicted with the help of computer analysis using the GeneFINDER program in PomBase (an ACEDB database) with additional predictions for the branch-acceptor sites supplied by the program Sp3splice. CAUTION: It is possible that for any individual CDS we may have underestimated or overestimated the number of introns/exons or we may not have chosen the correct splice donor/acceptor sites.
CDS are numbered using the following system eg SPBC35H2.01c. SP (S. pombe), B (chromosome 2), c25H2 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous.
The length in codons is given for each CDS.
IMPORTANT: This sequence MAY NOT be the entire insert of the

sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions.
cosmid c16c4 is overlapped at the 5' end by cosmid SPC05E4.
EM:AL033406.

FEATURES

```

source
    Location/Qualifiers
        1..44751
            /organism="Schizosaccharomyces pombe"
            /strain="972h-"
            /db_xref="taxon:4896"
            /chromosome="III"
            /clone="cosmid c16c4"
            /map="III"
        gene
            1..426
                /gene="SPCC16C4.19"
                /note="SPCC5E4.08"
        misc_feature
            1..1736
                /note="nominal overlap with cosmid SPC05E4, EM:AL033406 S. pombe chromosome 3"
            1..426
                /partial
                /gene="SPCC16C4.19"
                /note="SPCC16C4.19, len:>140, SIMILARITY: Saccharomyces cerevisiae, RMR-YEAST, mase mrp protein component sm1., (198 aa), fasta scores: opt: 142, E():0.0035, (26.0% identity in 146 aa)"
                /codon_start=1
                /label="SPCC16C4.19"
                /product="hypothetical protein"
                /protein_id="CA871193.1"
                /db_xref="GI:6855454"
                /translation="DPKTLHGSCGKSLFIFGRNLIKRTISERKARTDLKGEYSR LGNLEINLCVSYHEIPLGVPLRLHOSKOLRAKAIESSSTKLESKSAHNAVK QQRRLASGLNGIDPRKKKDEYAKSTSLSDFMSPY"
                join(1000..1307,1391..1656,1735..1796,1877..2581)
                /gene="SPCC16C4.01"
                join(1000..1307,1391..1656,1735..1796,1877..2581)
                /note="SPCC16C4.01"
                /gene="SPCC16C4.01"
                /note="SPCC16C4.01, len:446, SIMILARITY: Saccharomyces cerev istiae, Q03441, hypothetical 49.9 kd protein., (430 aa), fas ta scores: opt: 995, E():0, (34.8% identity in 376 aa)"
                /codon_start=1
                /label="SPCC16C4.01"
                /product="conserved hypothetical protein"
                /protein_id="CA20740.1"
                /db_xref="GI:3560255"
                /translation="MSNRIGPQSTKTAKLRLLPSTEEFDFRRDGTGREVYSQIPQ IEGSTAKRDAEHLGKRHREFLPRTAYCTDPRVDLRFQSRSSSHKTRKQDE CIYSPSYNNEETDLPLPTLESSRGTLRBSSQESLQIFESGLDRNQPFLREYFC FTYGVVLYNGITIDERRHRLRELGREIEKLEKIDMEVEEFNYITTLQPRIFNFI ALRDASNVMIRLSISHAIAOSVKISLFEELVNETIDATKDTQPMIAETGRVNLKREI MMAVGQFLIRININIGSVLDSPELMEPELEPYTAARSYLEINQVALINQVE VIGDLISLMEKQITHTHDESLMEIIVLGLVLLALFSLIVRLADGFLNDRFLVIN NSNKFLLPMLNLFYIGLEAVSITTEESFLFLHYIIIEEMGKKKIKIRPYLSS"
        misc_feature
            1308..1313
                /note="gtcgtt, splice donor sequence"
        misc_feature
            1376..1390
                /note="ctaacattataacag, splice branch and acceptor"
        misc_feature
            1657..1662
                /note="gtatgc, splice donor sequence"
        misc_feature
            1715..1734
                /note="ttaacattgtaatatttag, splice branch and acceptor"
        misc_feature
            1797..1802
                /note="gtatga, splice donor sequence"
        misc_feature
            1864..1876
                /note="ctaacagatatag, splice branch and acceptor"
        gene
            complement(join(3019..3212,3265..3292,3338..3349))
            /gene="SPCC16C4.20c"
            /note="SPCC16C4.20c"
        CDS
            /note="SPCC16C4.20c, len:77"

```

```

        misc_feature
            /codon_start=1
            /label="SPCC16C4.20c"
            /product="hypothetical protein"
            /protein_id="CA871194.1"
            /db_xref="GI:6855455"
            /translation="MDPVVNLQETEDVYRRKTAIRRRKSEIEKECBMAVKLYN VERDIORYRHEAQLSLKLEKLNKEKAGEY"
        intron
            /note="ctaacattacagatatag, splice branch and acceptor"
            complement(3213..3322)
            /note="confirmed intron"
            complement(3213..3264)
            /note="gtatgc, splice donor sequence"
            complement(3259..3264)
            /note="gtatgc, splice donor sequence"
            complement(3293..3304)
            /note="ttgaccgcgtag, splice branch and acceptor"
            complement(3332..3337)
            /note="gtatga, splice donor sequence"
            complement(4510..6156)
            /gene="SPCC16C4.02c"
            complement(4510..6156)
            /gene="SPCC16C4.02c"
            /note="SPCC16C4.02c, len:548"
        CDS
            /codon_start=1
            /label="SPCC16C4.02c"
            /product="hypothetical protein"
            /protein_id="CA20741.1"
            /db_xref="GI:3560256"
            /db_xref="SPRMBL:074447"
            /translation="MHIPPHLHGKGYRTISYEDLLSEDSYASEKISSEHYEVH FVYDKEDSNASGSRGSMELLENCSLHAODSTKPVSLMLALNDHPNLIKRC WERMMDKFLDRLRLSTHYEYVDVGSILAFCSGEALILSYEVKRRVSTLQCLKHY DCLTPIVCTLSNPKSAKYLAYTSFINPEPQAEILSNLYALNDVQYTRPIE OGIDRNGKMLKQCTFEFSNLSFRPVOVSYSIAIANIOPIMDAVERITDNKLS AYILSNLKAAGRPASIMPRDGMILIVIRGCSERIGSLGMLYKAGQGRKHTVYT VECYVILGLIRYLCEKNDLAORIEPDRFQLOSLTELSDTMDFLADAWNNK RNLASHVTVISAVATLCMLTEDDSQYAOASGLMDIFYLMSHWSNGIDAKWISY ALPMSLNKRVFFKAFDPDMKVYVDQFLKNDNDLGKSFNDYIISTWEDEDERL AQAIDPFILIOINSLVPOSIMNDIMQEPYKMLLESNF"
            join(6836..7236,7382..7508)
            /gene="SPCC16C4.03"
            join(6836..7236,7382..7508)
            /note="SPCC16C4.03"
            /note="SPCC16C4.03, len:175, SIMILARITY: Emeritella nidulans, Q42735, peptidyl-prolyl cis/trans isomerase., (176 aa), f asta scores: opt: 666, E():0, (57.5% identity in 174 aa)"
            /codon_start=1
            /label="SPCC16C4.03"
            /product="peptidyl-prolyl cis-trans isomerase."
            /protein_id="CA20742.1"
            /db_xref="GI:3560257"
            /db_xref="SPRMBL:074448"
            /translation="MSNTQSLPKRWIVKISRNRNPFYFNTTHESLMEPPAATDMAL KRFIANELEDVYTPPEASNSPKIRASHLIVKHRSRPSMKKEHTTRSEERKLAIE HKEOLKSGSVSHMDLAKMESDSSARRGELGEFGRDEMQKPFDAARALKGEISG VVETSSGFIHLIRHNA"
        misc_feature
            6851..6943
                /gene="SPCC16C4.03"
                /note="Match to Pf00397 WW_rsp5_WWP, WW domain Score 45.65"
                join(7025..7236,7382..7505)
                /gene="SPCC16C4.03"
                /note="Match to Pf00639 Rotamase, PPIC-type PPIASE domain. Score 138.03"
                /note="gtcgtt, splice donor sequence"
                7237..7242
                /note="gtatgc, splice donor sequence"
                7366..7381
                /note="ctaacattatgttag, splice branch and acceptor"
                complement(7505..11633)
            /note="possible gene free region, low complexity, SPC16C4.14c and SPC16C4.15c are predicted in this region, but have very low coding potential and no homology"

```



```

mRNA      RLAGVTLVAPVNYWYWRNLPINISTEGFNLQQRQDAVRAVAYAPWLYWNTQNMF
          PGSSVYNRQGVSPDKDIIKLGSRRKPHIAEVQOQIHESIINDMIVGFENMED
          PLEENPFLNRGSVHLMQODEMLVPTLQRYIADKLPLHLYHEVAAGGHHFPLAKG
          YVDEIVKTLAINH
          complement(join<18687..19284,20351..20547,20799..>21119))
gene      /gene="F1017.4"
          complement(18687..21119)
          /gene="F1017.4"
          /note="similar to unknown protein GB:AAD21437 from
[Arabidopsis thaliana]"
CDS       complement(join(18687..19284,20351..20547,20799..21119))
          /gene="F1017.4"
          /codon_start=1
          /product="unknown protein; 21119-18687"
          /protein_id="AAG52412.1"
          /db_xref="GI:12324910"
          /translation="MASDSSRSLSLKPPSRKSLFSPSVIIVIGIIVALTYSKLP
          OPRICGSSGPPITAPRIKLDGRYLAKHGLPREKANRIYFIHSGCCRDAAVA
          TILSPDLVELGYVWSPDPCESDPHSPRSLVSDIEIDLQSLGSKFYIG
          KSMGGAAMGCNLKXIPRILAGVTLVAPVNYWYWRNLPINISTEGFNLQQRQDAV
          RYAHYAPWLYWNTQNMFPKSSIANRDSLSQSDDIISKRYTRKPRHAEVROGI
          HESINRDMIVGFENMEDPLDNDPFLNNEGVHLMQODEMLVPVKLQRYLAHQLPW
          VHYHEVRSRSHFHFPTKGVVDNIVTTLTDTDTIQ"
mRNA      complement(join<21829..22626,22877..>23197))
          /gene="F1017.5"
          /codon_start=1
          /product="unknown protein; 23197-21829"
          /protein_id="AAG52399.1"
          /db_xref="GI:12324897"
          /translation="MTRDSRNSLSLKMPRKFLFSPSVIIVIGIIVALTYSKLP
          PRLICGSSGPPITAPRIKLDGRYLAKHGLPREKANRIYFIHSGCCRDAAVA
          TILSPDLVELGYVWSPDPCESDPHSPRSLVSDIEIDLQSLGSKFYIG
          YSMGGAAMGCNLKXIPRILAGVTLVAPVNYWYWRNLPINISTEGFNLQQRQDAV
          RYAHYAPWLYWNTQNMFPKSSIANRDSLSQSDDIISKRYTRKPRHAEVROGI
          HESINRDMIVGFENMEDPLDNDPFLNNEGVHLMQODEMLVPVKLQRYLAHQLPW
          VHYHEVRSRSHFHFPTKGVVDNIVTTLTDTDTIQ"
mRNA      join(<23489..23520,23750..23785,24176..24404,24499..24554)
          /gene="F1017.6"
          /codon_start=1
          /product="putative ribosomal protein; 23489-24540"
          /protein_id="AAG52401.1"
          /db_xref="GI:12324899"
          /translation="MYKRGGRGSRVRLYVGTLLGYKRSKNQYPTSLVQIEGVNTOE
          ENWYKGRMAYITVAKTKKNGSHYRCINGKVTTPHNGSGVYRAKRTSNLPPKSMGR
          VRFVWYPSNT"
mRNA      join(<25509..25911,25995..26563,26663..29047,
          29144..>29950)
          /gene="F1017.7"
          /codon_start=1
          /product="putative phosphoribosylformylglycinamide
          synthase; 25509-29950"
          /protein_id="AAG52403.1"
          /db_xref="GI:12324901"
          /translation="MLQRSSWSQMLGWSYRMTSRSLNRTKAVSLRCSAQPKNPKRAA
          13.8%; Score 17; DB 8; Length 50821;

```

```

Best Local Similarity 100.0%; Pred. No. 62;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 74 taattccacatgaaag 90
Db 8677 TAAGTTTCACATGTAAG 8661

RESULT 41
AC018376
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
2 (bases 1 to 59935)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bede,F.,
Boguslavskiy,L., Boukhalter,B., Brown,A., Castle,A., Colangelo,M.,
Collins,S., Collymore,A., Cooke,P., DeRubeis,K., Dewar,K.,
Domino,M., Doyle,M., Feneisior,J., Ferreira,P., Fitzhugh,M.,
Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B.,
Hearford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L.,
Karatas,A., Klein,J., Landers,T., Lechoczky,J., Lieu,C., Locke,K.,
Macdonald,P., Margulis,N., McEwan,P., McGurk,A., McKernan,K.,
Meldrim,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T.,
O'Donnell,P., Peterson,K., Pierre,N., Pollara,V., Riley,R.,
Rothman,D., Roy,A., Santos,R., Severy,P., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (10-DEC-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6554602.
All repeats were identified using RepeatMasker:
Smit,A.F.A. & Green,P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2801
Center clone name: 3_p_5
-----
* NOTE: This record contains 60 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 915: contig of 915 bp in length
* 916 1015: gap of 100 bp
* 1016 1921: contig of 906 bp in length
* 1922 2021: gap of 100 bp
* 2022 2930: contig of 909 bp in length
* 2931 3030: gap of 100 bp
* 3031 3907: contig of 877 bp in length

```

```

* 3908 4007: gap of 100 bp
* 4008 4914: contig of 907 bp in length
* 4915 5014: gap of 100 bp
* 5015 5914: contig of 900 bp in length
* 5915 6014: gap of 100 bp
* 6015 6932: contig of 918 bp in length
* 6933 7032: gap of 100 bp
* 7033 7950: contig of 918 bp in length
* 7951 8050: gap of 100 bp
* 8051 8949: contig of 899 bp in length
* 8950 9049: gap of 100 bp
* 9050 9934: contig of 885 bp in length
* 9935 10034: gap of 100 bp
* 10035 10935: contig of 901 bp in length
* 10936 11035: gap of 100 bp
* 11036 11948: contig of 913 bp in length
* 11949 12048: gap of 100 bp
* 12049 12963: contig of 915 bp in length
* 12964 13063: gap of 100 bp
* 13064 13966: contig of 903 bp in length
* 13967 14066: gap of 100 bp
* 14067 14999: contig of 933 bp in length
* 15000 15099: gap of 100 bp
* 15100 15984: contig of 885 bp in length
* 15985 16084: gap of 100 bp
* 16085 16956: contig of 872 bp in length
* 16957 17056: gap of 100 bp
* 17057 17940: contig of 884 bp in length
* 17941 18040: gap of 100 bp
* 18041 18964: contig of 924 bp in length
* 18965 19064: gap of 100 bp
* 19065 19995: contig of 931 bp in length
* 19996 20095: gap of 100 bp
* 20096 21002: contig of 907 bp in length
* 21003 21102: gap of 100 bp
* 21103 22002: contig of 900 bp in length
* 22003 22102: gap of 100 bp
* 22103 23023: contig of 921 bp in length
* 23024 23123: gap of 100 bp
* 23124 23997: contig of 874 bp in length
* 23998 24097: gap of 100 bp
* 24098 25011: contig of 914 bp in length
* 25012 25111: gap of 100 bp
* 25112 26014: contig of 903 bp in length
* 26015 26114: gap of 100 bp
* 26115 27008: contig of 894 bp in length
* 27009 27108: gap of 100 bp
* 27109 27988: contig of 880 bp in length
* 27989 28088: gap of 100 bp
* 28089 29008: contig of 920 bp in length
* 29009 29108: gap of 100 bp
* 29109 30007: contig of 899 bp in length
* 30008 30107: gap of 100 bp
* 30108 31001: contig of 894 bp in length
* 31002 31101: gap of 100 bp
* 31102 31983: contig of 882 bp in length
* 31984 32083: gap of 100 bp
* 32084 32967: contig of 884 bp in length
* 32968 33067: gap of 100 bp
* 33068 33986: contig of 919 bp in length
* 33987 34086: gap of 100 bp
* 34087 34995: contig of 909 bp in length
* 34996 35095: gap of 100 bp
* 35096 36028: contig of 933 bp in length
* 36029 36128: gap of 100 bp
* 36129 36995: contig of 867 bp in length
* 36996 37095: gap of 100 bp
* 37096 38018: contig of 923 bp in length
* 38019 38118: gap of 100 bp
* 38119 39019: contig of 901 bp in length
* 39020 39119: gap of 100 bp
* 39120 40012: contig of 893 bp in length
* 40013 40112: gap of 100 bp

```

```

* 40113 41016: contig of 904 bp in length
* 41017 41116: gap of 100 bp
* 41117 42005: contig of 889 bp in length
* 42006 42105: gap of 100 bp
* 42106 43055: contig of 950 bp in length
* 43056 43155: gap of 100 bp
* 43156 44088: contig of 933 bp in length
* 44089 44188: gap of 100 bp
* 44189 45107: contig of 919 bp in length
* 45108 45207: gap of 100 bp
* 45208 46128: contig of 921 bp in length
* 46129 46228: gap of 100 bp
* 46229 47142: contig of 914 bp in length
* 47143 47242: gap of 100 bp
* 47243 48115: contig of 873 bp in length
* 48116 48215: gap of 100 bp
* 48216 48936: contig of 721 bp in length
* 48937 49036: gap of 100 bp
* 49037 49922: contig of 886 bp in length
* 49923 50022: gap of 100 bp
* 50023 50946: contig of 924 bp in length
* 50947 51046: gap of 100 bp
* 51047 51950: contig of 904 bp in length
* 51951 52050: gap of 100 bp
* 52051 52962: contig of 912 bp in length
* 52963 53062: gap of 100 bp
* 53063 53922: contig of 860 bp in length
* 53923 54022: gap of 100 bp
* 54023 54959: contig of 937 bp in length
* 54960 55059: gap of 100 bp
* 55060 55924: contig of 865 bp in length
* 55925 56024: gap of 100 bp
* 56025 56942: contig of 918 bp in length
* 56943 57042: gap of 100 bp
* 57043 57936: contig of 894 bp in length
* 57937 58036: gap of 100 bp
* 58037 58933: contig of 897 bp in length
* 58934 59033: gap of 100 bp
* 59034 59935: contig of 902 bp in length.

```

FEATURES
source 1. 59935
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-3p5"
/clone_lib="RPCI-11 Human Male BAC"

BASE COUNT 16503 a 10402 c 9763 g 17088 t 6179 others
ORIGIN

Query Match 13.8%; Score 17; DB 2; Length 59935;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 aggtattgtaaatgga 18
DB 5750 AGGTATTGTAATGGA 5766
|||||

RESULT 42

LOCUS AP000381 63604 bp DNA 27-DEC-2000
DEFINITION Arabidopsis thaliana genomic DNA, chromosome 3, TAC clone:K17E12.
ACCESSION AP000381 BA000014
VERSION AP000381.1 GI:5672506

KEYWORDS Arabidopsis thaliana (strain:Columbia) DNA, clone_lib:Mitsui TAC
SOURCE Clone:K17E12.

ORGANISM

Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.

REFERENCE 1 (sites)
AUTHORS Kaneko,T., Katoh,T., Sato,S., Nakamura,A., Asamizu,E. and Tabata,S.


```

34154..34241,34515..34595,34685..34794,35165..35234,
35378..35520,35718..36272)
/contig_id="gb|AA01563.1
gene_id="K17E12.8
similar to unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB02121.1"
/db_xref="GI:9294219"
/translation="WESAAFPGGYRRNTPEAPESESGSSQAIDTEVVASNSSTP
ARKCINLNSNDDEPYGVROVISLYNMSKSEKDLIYRKILEQRIYVKNELQRM
NPAVSTSDRYGSESGKISSRVSNSKPSDPAVSGSKRHQNDTSRCGRMNSTGSK
FESSKRTMTSTPNTITLAKOCOTILRLKNSHPHNVQAPDYVKINIPDTITIKPM
DGTIVKKNLASCYSSPHEFADVRLFTNMTYNPQGHVIMGDIKSLFEARKMT
IKKLPCCSMQTLPAVTLTPNDEKKAISVPPAKRKMASPVRESVPEPKPLMTEVE
HRLRGKLESLDELPAHITIDFLKNSNGGAEDEIEIDVLSDEVLYPLRLNLD
EYIIONKEAKOTPECEIEILINGSRNSSLGNGEMADYEIDVLSDEVLYPLRLNLD
GSSGSDODAKPMVGDSKSLVNLQMPETANSQORDENTRIDDLVGSOSTALQOM
DICSQOKLSSDSDSGHESNILETPASSSEKRYRAALKNRADIIILAKRKPPIQNGI
KQDPERLKERELVLOKKERKARLOAEADARKQAEAEAAAEAAAEAKRRELE
REARQALLKMEKYEIINNSRFLQEMISSAPQOLPSAETSPEPRLDLSGFN
LRGNSPDLQGLYMKODDEEPEAPAVKPDPTSTERPDLALGSFKLGCPNLEQIG
LYMKODDGESEPEAPAVKPNPNTSLERPVDAGSFNLKGSNPLQGLYMKODDDEE
PEAPAVPLANDVEGEID"
37184..37933
/contig_id="K17E12.9
gene_id="K17E12.9
unknown protein"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB02121.1"
/db_xref="GI:9294220"
/translation="MEPTIYSKDLPLFESIFIFVYLLGYLFIKKWTPETRLASSCL
ISLHGSVAVVATNALISDPNGSFSSVNTQSONSLIDSSAFDLADIVLHAFPSA
GSDSLFAHHLAVLEFVLCRYVAVAGACALLVVAATSCQMTWLADRGDA
PLAVSLHREVTVPEFYASVCGCVLAPLLIKVMTFVYSGADVDIPRMVWSWTVI
VTAVVSLIMINIVLFEORRYSKRTKRV"
join(40773..41084,41393..41920)
/contig_id="K17E12.10"
gene_id="K17E12.10"
/codon_start=1
/evidence=not_experimental
Query Match 13.8%; Score 17; DB 8; Length 63604;
Best Local Similarity 100.0%; Pred. No. 61;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 95 tactatctcaattta 111
|||||
Db 34800 TACTTATCTTCATTTA 34816

```

RESULT 43
AC018413/C
LOCUS AC018413 65857 bp DNA HTG 12-JAN-2001
DEFINITION Homo sapiens clone Rpl11-17M16, LOW-PASS SEQUENCE SAMPLING.
AC018413
VERSION AC018413.3 GI:12084068
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 65857)
TITLE Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens chromosome, clone Rpl11-17M16
Unpublished
2 (bases 1 to 65857)
TITLE Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Andersen,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Boukhalter,B., Brown,A., Castle,A., Colangelo,M.,
Collins,S., Collymore,A., Cooke,P., Dearellano,K., Dewar,K.,
Domino,M., Doyle,M., Fenestor,J., Ferreira,P., Fitzhugh,W.,
Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B.,

Heaford A., Horton, L., Howland, J.C., Johnson, R., Jones, C., Kamp, L.,
Karatas, A., Klein, J., Landers, T., Lenochsky, J., Lieu, C., Locke, K.,
Macdonald, P., Margulis, N., McEwan, P., McGurk, A., McKernan, K.,
Meldrum, J., Morrow, J., Naylor, J., Norman, C.H., O'Connor, T.,
O'Donnell, P., Peterson, K., Pierre, N., Pollara, V., Riley, R.,
Rothman, D., Roy, A., Santos, R., Severy, P., Strange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
Zimmer, A. and Zody, M.
Direct Submission
Submitted (10-DEC-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jan 12, 2001 this sequence version replaced gi:9123871.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/xw/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3661
Center clone name: 17_M16

* NOTE: This record contains 80 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
725 824: gap of 100 bp
825 1548: contig of 724 bp in length
1549 1648: gap of 100 bp
1649 2375: contig of 727 bp in length
2376 2475: gap of 100 bp
2476 3224: contig of 749 bp in length
3225 3324: gap of 100 bp
3325 4044: contig of 720 bp in length
4045 4144: gap of 100 bp
4145 4864: contig of 720 bp in length
4865 4964: gap of 100 bp
4965 5687: contig of 723 bp in length
5688 5787: gap of 100 bp
5788 6498: contig of 711 bp in length
6499 6598: gap of 100 bp
6599 7310: contig of 712 bp in length
7311 7410: gap of 100 bp
7411 8149: contig of 739 bp in length
8150 8249: gap of 100 bp
8250 8993: contig of 744 bp in length
8994 9093: gap of 100 bp
9094 9822: contig of 729 bp in length
9823 9922: gap of 100 bp
9923 10654: contig of 732 bp in length
10655 10754: gap of 100 bp
10755 11498: contig of 744 bp in length
11499 11598: gap of 100 bp
11599 12337: contig of 739 bp in length
12338 12437: gap of 100 bp
12438 13134: contig of 697 bp in length
13135 13234: gap of 100 bp
13235 13951: contig of 717 bp in length
13952 14051: gap of 100 bp
14052 14777: contig of 726 bp in length
14778 14877: gap of 100 bp
14878 15594: contig of 717 bp in length


```

* 15595 15694: gap of 100 bp
* 15695 16405: contig of 711 bp in length
* 16406 16505: gap of 100 bp
* 16506 17186: contig of 681 bp in length
* 17187 17286: gap of 100 bp
* 17287 18018: contig of 732 bp in length
* 18019 18118: gap of 100 bp
* 18119 18843: contig of 725 bp in length
* 18844 18943: gap of 100 bp
* 18944 19668: contig of 725 bp in length
* 19669 19768: gap of 100 bp
* 19769 20505: contig of 737 bp in length
* 20506 20605: gap of 100 bp
* 20606 21340: contig of 735 bp in length
* 21341 21440: gap of 100 bp
* 21441 22150: contig of 710 bp in length
* 22151 22250: gap of 100 bp
* 22251 22981: contig of 731 bp in length
* 22982 23081: gap of 100 bp
* 23082 23812: contig of 731 bp in length
* 23813 23912: gap of 100 bp
* 23913 24631: contig of 719 bp in length
* 24632 24731: gap of 100 bp
* 24732 25459: contig of 728 bp in length
* 25460 25559: gap of 100 bp
* 25560 26221: contig of 662 bp in length
* 26222 26321: gap of 100 bp
* 26322 27049: contig of 728 bp in length
* 27050 27149: gap of 100 bp
* 27150 27886: contig of 737 bp in length
* 27887 27986: gap of 100 bp
* 27987 28708: contig of 722 bp in length
* 28709 28808: gap of 100 bp
* 28809 29536: contig of 728 bp in length
* 29537 29636: gap of 100 bp
* 29637 30363: contig of 727 bp in length
* 30364 30463: gap of 100 bp
* 30464 31182: contig of 719 bp in length
* 31183 31282: gap of 100 bp
* 31283 31991: contig of 709 bp in length
* 31992 32091: gap of 100 bp
* 32092 32811: contig of 720 bp in length
* 32812 32911: gap of 100 bp
* 32912 33622: contig of 711 bp in length
* 33623 33722: gap of 100 bp
* 33723 34412: contig of 690 bp in length
* 34413 34512: gap of 100 bp
* 34513 35248: contig of 736 bp in length
* 35249 35348: gap of 100 bp
* 35349 36092: contig of 744 bp in length
* 36093 36192: gap of 100 bp
* 36193 36909: contig of 717 bp in length
* 36910 37009: gap of 100 bp
* 37010 37740: contig of 731 bp in length
* 37741 37840: gap of 100 bp
* 37841 38573: contig of 733 bp in length
* 38574 38673: gap of 100 bp
* 38674 39380: contig of 707 bp in length
* 39381 39480: gap of 100 bp
* 39481 40201: contig of 721 bp in length
* 40202 40301: gap of 100 bp
* 40302 41017: contig of 716 bp in length
* 41018 41117: gap of 100 bp
* 41118 41820: contig of 703 bp in length
* 41821 41920: gap of 100 bp
* 41921 42648: contig of 728 bp in length
* 42649 42748: gap of 100 bp
* 42749 43483: contig of 735 bp in length
* 43484 43583: gap of 100 bp
* 43584 44321: contig of 738 bp in length
* 44322 44421: gap of 100 bp
* 44422 45147: contig of 726 bp in length
* 45148 45247: gap of 100 bp

```

```

* 45248 45965: contig of 718 bp in length
* 45966 46065: gap of 100 bp
* 46066 46816: contig of 751 bp in length
* 46817 46916: gap of 100 bp
* 46917 47574: contig of 658 bp in length
* 47575 47674: gap of 100 bp
* 47675 48405: contig of 731 bp in length
* 48406 48505: gap of 100 bp
* 48506 49200: contig of 695 bp in length
* 49201 49300: gap of 100 bp
* 49301 50023: contig of 723 bp in length
* 50024 50123: gap of 100 bp
* 50124 50855: contig of 732 bp in length
* 50856 50955: gap of 100 bp
* 50956 51656: contig of 701 bp in length
* 51657 51756: gap of 100 bp
* 51757 52503: contig of 747 bp in length
* 52504 52603: gap of 100 bp
* 52604 53342: contig of 739 bp in length
* 53343 53442: gap of 100 bp
* 53443 54212: contig of 770 bp in length
* 54213 54312: gap of 100 bp
* 54313 55040: contig of 728 bp in length
* 55041 55140: gap of 100 bp
* 55141 55860: contig of 720 bp in length
* 55861 55960: gap of 100 bp
* 55961 56702: contig of 742 bp in length
* 56703 56802: gap of 100 bp
* 56803 57556: contig of 754 bp in length
* 57557 57656: gap of 100 bp
* 57657 58386: contig of 730 bp in length
* 58387 58486: gap of 100 bp
* 58487 59224: contig of 738 bp in length
* 59225 59324: gap of 100 bp

```

```

Query Match      13.8%; Score 17; DB 2; Length 65857;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      3  ggtatgttaagtggag 19
Db      4032  GGTATGTGTAATGTGAG 4016

```

```

RESULT  44
LOCUS   AC023633/c      65998 bp      DNA      HTG      13-JUL-2000
DEFINITION Homo sapiens clone RP11-287J24, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC023633
VERSION   AC023633.2 GI:9144869
KEYWORDS  HTG; HTGS_PHASE0.
SOURCE   human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 65998)
Birken,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-287J24
Unpublished
REFERENCE 2 (bases 1 to 65998)
AUTHORS  Anderson,S., Baldwin,J., Barna,N., Beka,F., Boguslavsky,L.,
Bouknight,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
Dearellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fensterlo,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Largocque,K., Lehoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., Macdonald,P., Margus,N., McCarthy,M.,
McEwan,P., McGurk,A., McKernan,K., McPheters,R., Meldrim,J.,
Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,

```

TITLE
JOURNAL
COMMENT

Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Olivar,T.M.,
Peterson,K., Pierre,N., Pisanl,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severy,P., Spencer,B., Strange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A.,
Travers,M., Trigglio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.

Direct Submission

Submitted (16-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6980301.
All repeats were identified using RepeatMasker:

Smit,A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Genome Center

Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Center project name: L5205

Center clone name: 287_J_24

***** NOTE: This record contains 75 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 769: contig of 769 bp in length
* 770 869: gap of 100 bp
* 870 1650: contig of 781 bp in length
* 1651 1750: gap of 100 bp
* 1751 2538: contig of 788 bp in length
* 2539 2638: gap of 100 bp
* 2639 3437: contig of 799 bp in length
* 3438 3537: gap of 100 bp
* 3538 4338: contig of 801 bp in length
* 4339 4438: gap of 100 bp
* 4439 5234: contig of 796 bp in length
* 5235 5334: gap of 100 bp
* 5335 6108: contig of 774 bp in length
* 6109 6208: gap of 100 bp
* 6209 6972: contig of 764 bp in length
* 6973 7072: gap of 100 bp
* 7073 7701: contig of 629 bp in length
* 7702 7801: gap of 100 bp
* 7802 8604: contig of 803 bp in length
* 8605 8704: gap of 100 bp
* 8705 9481: contig of 777 bp in length
* 9482 9581: gap of 100 bp
* 9582 10290: contig of 709 bp in length
* 10291 10390: gap of 100 bp
* 10391 11174: contig of 784 bp in length
* 11175 11274: gap of 100 bp
* 11275 12066: contig of 792 bp in length
* 12067 12166: gap of 100 bp
* 12167 12931: contig of 765 bp in length
* 12932 13031: gap of 100 bp
* 13032 13781: contig of 750 bp in length
* 13782 13881: gap of 100 bp
* 13882 14655: contig of 774 bp in length
* 14656 14755: gap of 100 bp
* 14756 15591: contig of 836 bp in length
* 15592 15691: gap of 100 bp
* 15692 16473: contig of 782 bp in length
* 16474 16573: gap of 100 bp

16574 17358: contig of 785 bp in length
* 17359 17458: gap of 100 bp
* 17459 18266: contig of 808 bp in length
* 18267 18366: gap of 100 bp
* 18367 19147: contig of 781 bp in length
* 19148 19247: gap of 100 bp
* 19248 20033: contig of 786 bp in length
* 20034 20133: gap of 100 bp
* 20134 20934: contig of 801 bp in length
* 20935 21034: gap of 100 bp
* 21035 21822: contig of 788 bp in length
* 21823 21922: gap of 100 bp
* 21923 22720: contig of 798 bp in length
* 22721 22820: gap of 100 bp
* 22821 23610: contig of 790 bp in length
* 23611 23710: gap of 100 bp
* 23711 24480: contig of 770 bp in length
* 24481 24580: gap of 100 bp
* 24581 25378: contig of 796 bp in length
* 25379 25478: gap of 100 bp
* 25479 26278: contig of 800 bp in length
* 26279 26378: gap of 100 bp
* 26379 27162: contig of 784 bp in length
* 27163 27262: gap of 100 bp
* 27263 28041: contig of 779 bp in length
* 28042 28141: gap of 100 bp
* 28142 28942: contig of 801 bp in length
* 28943 29042: gap of 100 bp
* 29043 29840: contig of 798 bp in length
* 29841 29940: gap of 100 bp
* 29941 30733: contig of 793 bp in length
* 30734 30833: gap of 100 bp
* 30834 31632: contig of 799 bp in length
* 31633 31732: gap of 100 bp
* 31733 32530: contig of 798 bp in length
* 32531 32630: gap of 100 bp
* 32631 33422: contig of 792 bp in length
* 33423 33522: gap of 100 bp
* 33523 34280: contig of 758 bp in length
* 34281 34380: gap of 100 bp
* 34381 35184: contig of 804 bp in length
* 35185 35284: gap of 100 bp
* 35285 36073: contig of 789 bp in length
* 36074 36173: gap of 100 bp
* 36174 36967: contig of 794 bp in length
* 36968 37067: gap of 100 bp
* 37068 37838: contig of 771 bp in length
* 37839 37938: gap of 100 bp
* 37939 38703: contig of 765 bp in length
* 38704 38803: gap of 100 bp
* 38804 39549: contig of 746 bp in length
* 39550 39649: gap of 100 bp
* 39650 40409: contig of 760 bp in length
* 40410 40509: gap of 100 bp
* 40510 41284: contig of 775 bp in length
* 41285 41384: gap of 100 bp
* 41385 42161: contig of 777 bp in length
* 42162 42261: gap of 100 bp
* 42262 43048: contig of 787 bp in length
* 43049 43148: gap of 100 bp
* 43149 43910: contig of 762 bp in length
* 43911 44010: gap of 100 bp
* 44011 44774: contig of 764 bp in length
* 44775 44874: gap of 100 bp
* 44875 45644: contig of 770 bp in length
* 45645 45744: gap of 100 bp
* 45745 46598: contig of 854 bp in length
* 46599 46698: gap of 100 bp
* 46699 47466: contig of 768 bp in length
* 47467 47566: gap of 100 bp
* 47567 48338: contig of 772 bp in length
* 48339 48438: gap of 100 bp
* 48439 49230: contig of 792 bp in length

```

* 49331 49330: gap of 100 bp
* 49331 50117: contig of 787 bp in length
* 50118 50217: gap of 100 bp
* 50218 51031: contig of 814 bp in length
* 51032 51131: gap of 100 bp
* 51132 51922: contig of 791 bp in length
* 51923 52022: gap of 100 bp
* 52023 52790: contig of 768 bp in length
* 52791 52890: gap of 100 bp
* 52891 53672: contig of 782 bp in length
* 53673 53772: gap of 100 bp
* 53773 54548: contig of 776 bp in length
* 54549 54648: gap of 100 bp
* 54649 55422: contig of 774 bp in length
* 55423 55522: gap of 100 bp
* 55523 56305: contig of 783 bp in length
* 56306 56405: gap of 100 bp
* 56406 57201: contig of 796 bp in length
* 57202 57301: gap of 100 bp
* 57302 58058: contig of 757 bp in length
* 58059 58158: gap of 100 bp
* 58159 58936: contig of 778 bp in length
* 58937 59036: gap of 100 bp
* 59037 59815: contig of 779 bp in length
* 59816 59915: gap of 100 bp
* 59916 60712: contig of 797 bp in length
* 60713 60812: gap of 100 bp

```

```

Query Match      13.88; Score 17; DB 2; Length 65998;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy 107 attattccacacaca 123
Db 5831 ATTATTCCACACACA 5815

```

```

RESULT 45
AC027574/c 70677 bp DNA HTG 30-MAR-2000
LOCUS Homo sapiens chromosome 15 clone RP11-149D17 map 15, LOW-PASS
DEFINITION AC027574
ACCESSION AC027574 GI:7342319
VERSION AC027574.1
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 70677)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens chromosome 15, clone RP11-149D17
REFERENCE 2 (bases 1 to 70677)
AUTHORS Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavskiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campione,A., Castle,A., Choepel,T., Colangelo,M., Collins,S.,
Collymore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W.,
Galaian,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Lakoque,K., Lamazares,R., Landers,T., Lehocsky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Melidim,J., Meneses,L., Mihova,T., Miranda,C., Mlenka,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Poltara,V., Raymond,C., Riley,K., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Strange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tessaye,S., Theodore,J., Tirrell,A., Travers,M., Trigglio,J.,

```

TITLE JOURNAL COMMENT

Vassiliev,R., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit,A.F. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L8988
Center clone name: 149_D_17

* NOTE: This record contains 80 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```

1
* 773 872: contig of 772 bp in length
* 773 872: gap of 100 bp
* 873 1670: contig of 798 bp in length
* 1671 1770: gap of 100 bp
* 1771 2581: contig of 811 bp in length
* 2582 2681: gap of 100 bp
* 2682 3489: contig of 808 bp in length
* 3490 3589: gap of 100 bp
* 3590 4393: contig of 804 bp in length
* 4394 4493: gap of 100 bp
* 4494 5286: contig of 793 bp in length
* 5287 5386: gap of 100 bp
* 5387 6176: contig of 790 bp in length
* 6177 6276: gap of 100 bp
* 6277 7065: contig of 789 bp in length
* 7066 7165: gap of 100 bp
* 7166 7959: contig of 794 bp in length
* 7960 8059: gap of 100 bp
* 8060 8837: contig of 778 bp in length
* 8838 8937: gap of 100 bp
* 8938 9734: contig of 797 bp in length
* 9735 9834: gap of 100 bp
* 9835 10632: contig of 798 bp in length
* 10633 10732: gap of 100 bp
* 10733 11499: contig of 767 bp in length
* 11500 11599: gap of 100 bp
* 11600 12397: contig of 798 bp in length
* 12398 12497: gap of 100 bp
* 12498 13274: contig of 777 bp in length
* 13275 13374: gap of 100 bp
* 13375 14156: contig of 782 bp in length
* 14157 14256: gap of 100 bp
* 14257 15049: contig of 793 bp in length
* 15050 15149: gap of 100 bp
* 15150 15927: contig of 778 bp in length
* 15928 16027: gap of 100 bp
* 16028 16813: contig of 786 bp in length
* 16814 16913: gap of 100 bp
* 16914 17695: contig of 782 bp in length
* 17696 17795: gap of 100 bp
* 17796 18594: contig of 799 bp in length
* 18595 18694: gap of 100 bp
* 18695 19501: contig of 807 bp in length
* 19502 19601: gap of 100 bp
* 19602 20369: contig of 768 bp in length

```

* 20370 20469: gap of 100 bp
* 20470 21294: contig of 825 bp in length
* 21295 21394: gap of 100 bp
* 21395 22195: contig of 801 bp in length
* 2196 22295: gap of 100 bp
* 22296 23026: contig of 731 bp in length
* 23027 23126: gap of 100 bp
* 23127 23864: contig of 738 bp in length
* 23865 23964: gap of 100 bp
* 23965 24741: contig of 777 bp in length
* 24742 25622: contig of 781 bp in length
* 25623 25722: gap of 100 bp
* 25723 26497: contig of 775 bp in length
* 26498 26597: gap of 100 bp
* 26598 27257: contig of 660 bp in length
* 27258 27357: gap of 100 bp
* 27358 28156: contig of 799 bp in length
* 28157 28256: gap of 100 bp
* 28257 29034: contig of 778 bp in length
* 29035 29134: gap of 100 bp
* 29135 29924: contig of 790 bp in length
* 29925 30024: gap of 100 bp
* 30025 30801: contig of 777 bp in length
* 30802 30901: gap of 100 bp
* 30902 31673: contig of 772 bp in length
* 31674 31773: gap of 100 bp
* 31774 32558: contig of 785 bp in length
* 32559 32658: gap of 100 bp
* 32659 33431: contig of 773 bp in length
* 33432 33531: gap of 100 bp
* 33532 34318: contig of 787 bp in length
* 34319 34418: gap of 100 bp
* 34419 35198: contig of 780 bp in length
* 35199 35298: gap of 100 bp
* 35299 36068: contig of 770 bp in length
* 36069 36168: gap of 100 bp
* 36169 36970: contig of 802 bp in length
* 36971 37070: gap of 100 bp
* 37071 37861: contig of 791 bp in length
* 37862 37961: gap of 100 bp
* 37962 38739: contig of 778 bp in length
* 38740 38839: gap of 100 bp
* 38840 39633: contig of 794 bp in length
* 39634 39733: gap of 100 bp
* 39734 40532: contig of 799 bp in length
* 40533 40632: gap of 100 bp
* 40633 41403: contig of 771 bp in length
* 41404 41503: gap of 100 bp
* 41504 42297: contig of 794 bp in length
* 42298 42397: gap of 100 bp
* 42398 43186: contig of 789 bp in length
* 43187 43286: gap of 100 bp
* 43287 44087: contig of 801 bp in length
* 44088 44187: gap of 100 bp
* 44188 44929: contig of 742 bp in length
* 44930 45029: gap of 100 bp
* 45030 45844: contig of 815 bp in length
* 45845 45944: gap of 100 bp
* 45945 46731: contig of 787 bp in length
* 46732 46831: gap of 100 bp
* 46832 47632: contig of 801 bp in length
* 47633 47732: gap of 100 bp
* 47733 48515: contig of 783 bp in length
* 48516 48615: gap of 100 bp
* 48616 49377: contig of 762 bp in length
* 49378 49477: gap of 100 bp
* 49478 50263: contig of 786 bp in length
* 50264 50363: gap of 100 bp
* 50364 51145: contig of 782 bp in length
* 51146 51245: gap of 100 bp
* 51246 52018: contig of 773 bp in length
* 52019 52118: gap of 100 bp

* 52119 52915: contig of 797 bp in length
* 52916 53015: gap of 100 bp
* 53016 53823: contig of 808 bp in length
* 53824 53923: gap of 100 bp
* 53924 54700: contig of 777 bp in length
* 54701 54800: gap of 100 bp
* 54801 55599: contig of 799 bp in length
* 55600 55699: gap of 100 bp
* 55700 56493: contig of 794 bp in length
* 56494 56593: gap of 100 bp
* 56594 57385: contig of 792 bp in length
* 57386 57485: gap of 100 bp
* 57486 58276: contig of 791 bp in length
* 58277 58376: gap of 100 bp
* 58377 59163: contig of 787 bp in length
* 59164 59263: gap of 100 bp
* 59264 60049: contig of 786 bp in length
* 60050 60149: gap of 100 bp
* 60150 60921: contig of 772 bp in length
* 60922 61021: gap of 100 bp

Query Match 13.8%; Score 17; DB 2; Length 70677;
Best Local Similarity 100.0%; Pred. No. 60;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 107 attatccacacaca 123
Db 32737 ATTATTCACACACACA 32721

Search completed: March 19, 2002, 11:05:47
Job time: 4021 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2002, 10:18:46 ; Search time 101.49 Seconds

(without alignments)
1039.029 Million cell updates/sec

Title: US-09-684-016-48411

Perfect score: 123

Sequence: 1 caggtatgttaatgtgaga.....tcaatttaccacacaca 123

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N.Geneseq-1101:*

- 1: /SIDS2/gcgcdata/geneseq/geneseq/NA1980.DAT:*
- 2: /SIDS2/gcgcdata/geneseq/geneseq/NA1981.DAT:*
- 3: /SIDS2/gcgcdata/geneseq/geneseq/NA1982.DAT:*
- 4: /SIDS2/gcgcdata/geneseq/geneseq/NA1983.DAT:*
- 5: /SIDS2/gcgcdata/geneseq/geneseq/NA1984.DAT:*
- 6: /SIDS2/gcgcdata/geneseq/geneseq/NA1985.DAT:*
- 7: /SIDS2/gcgcdata/geneseq/geneseq/NA1986.DAT:*
- 8: /SIDS2/gcgcdata/geneseq/geneseq/NA1987.DAT:*
- 9: /SIDS2/gcgcdata/geneseq/geneseq/NA1988.DAT:*
- 10: /SIDS2/gcgcdata/geneseq/geneseq/NA1989.DAT:*
- 11: /SIDS2/gcgcdata/geneseq/geneseq/NA1990.DAT:*
- 12: /SIDS2/gcgcdata/geneseq/geneseq/NA1991.DAT:*
- 13: /SIDS2/gcgcdata/geneseq/geneseq/NA1992.DAT:*
- 14: /SIDS2/gcgcdata/geneseq/geneseq/NA1993.DAT:*
- 15: /SIDS2/gcgcdata/geneseq/geneseq/NA1994.DAT:*
- 16: /SIDS2/gcgcdata/geneseq/geneseq/NA1995.DAT:*
- 17: /SIDS2/gcgcdata/geneseq/geneseq/NA1996.DAT:*
- 18: /SIDS2/gcgcdata/geneseq/geneseq/NA1997.DAT:*
- 19: /SIDS2/gcgcdata/geneseq/geneseq/NA1998.DAT:*
- 20: /SIDS2/gcgcdata/geneseq/geneseq/NA1999.DAT:*
- 21: /SIDS2/gcgcdata/geneseq/geneseq/NA2000.DAT:*
- 22: /SIDS2/gcgcdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	18	14.6	464	21	AAAC98121
2	18	14.6	1303	21	AAH29882
3	18	14.6	9941	21	AAH81461
4	18	14.6	349980	21	AAH21607
5	18	14.6	1437668	21	AAH81490
6	17	13.8	518	22	AAH29434
7	16	13.0	446	22	AAH03888
8	16	13.0	459	22	AAH14349
9	16	13.0	459	22	AAH35723
10	16	13.0	459	22	AAH04179
11	16	13.0	852	20	AAH34755

C 12	16	13.0	1278	21	AAA38323	Human angiotensino
C 13	16	13.0	1278	21	AAA38324	Human angiotensino
C 14	16	13.0	1491	14	AAH47075	21BA/rhoptery anti
C 15	16	13.0	2732	21	AAH98971	Human pancreatic c
C 16	16	13.0	4201	21	AAH78084	Human cancer assoc
C 17	16	13.0	4201	22	AAH33277	Human cancer cancer
C 18	16	13.0	5308	22	AAH91600	Human angiotensino
C 19	16	13.0	5308	22	AAH91670	Human angiotensino
C 20	16	13.0	5308	22	AAH91671	Human angiotensino
C 21	16	13.0	5308	22	AAH91672	Human angiotensino
C 22	16	13.0	5308	22	AAH91674	Human angiotensino
C 23	16	13.0	5308	22	AAH91675	Human angiotensino
C 24	16	13.0	5308	22	AAH91676	Human angiotensino
C 25	16	13.0	5308	22	AAH91677	Human angiotensino
C 26	16	13.0	50000	20	AAH23517	Human kidney amino
C 27	15	12.2	20	22	AAH91627	Human angiotensino
C 28	15	12.2	96	19	AAH72858	Human angiotensino
C 29	15	12.2	96	21	AAH65347	Human angiotensino
C 30	15	12.2	101	19	AAH72819	Human angiotensino
C 31	15	12.2	101	21	AAH65362	Human angiotensino
C 32	15	12.2	201	21	AAH36289	Mechanical stress
C 33	15	12.2	203	21	AAH13255	Human secreted pro
C 34	15	12.2	280	21	AAH01492	Human colon cancer
C 35	15	12.2	300	20	AAH21487	Human gene express
C 36	15	12.2	303	22	AAH67192	C glutamicum codin
C 37	15	12.2	309	21	AAH74896	Human OREX ORF451
C 38	15	12.2	319	20	AAH40689	Human secreted pro
C 39	15	12.2	357	21	AAH22215	Human secreted pro
C 40	15	12.2	377	22	AAH66728	Novel human polynu
C 41	15	12.2	391	22	AAH66301	Novel human polynu
C 42	15	12.2	412	22	AAH66405	Novel human polynu
C 43	15	12.2	434	21	AAH01762	Human secreted pro
C 44	15	12.2	556	21	AAH10271	Fusarium venenatum
C 45	15	12.2	567	21	AAH87998	Human CLAP-2 geno

ALIGNMENTS

RESULT 1	
AAH98121	standard; cDNA; 464 BP.
AC	
XX	AAH98121;
DF	09-MAR-2001 (first entry)
XX	
DE	Human colon cancer antigen nucleotide sequence SEQ ID NO:131.
XX	
KW	Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW	identification; cytosolic; cardiovascular; neuroprotective; vulnerrary;
KW	immunomodulatory; muscular; gynaecological; gastrointestinal;
KW	nephrotropic; antineoplastic; antibacterial; gene therapy; wound;
KW	neural disorder; immune system disorder; muscular disorder;
KW	reproductive disorder; gastrointestinal disorder; renal disorder;
KW	infectious disease; cardiovascular disorder; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200055351-A1.
XX	
PD	21-SEP-2000.
XX	
PF	08-MAR-2000; 2000MO-US05883.
XX	
PR	12-MAR-1999; 99US-0124270.
XX	
PA	(HUMA-) HUMAN GENOME SCI INC.
XX	
PI	Rosen CA, Ruben SM;
XX	
DR	WPI; 2000-587534/55.
DR	P-PDB; AAB53364.

XX Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
XX
PS Claim 1; Page 565; 2104pp; English.
XX
CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardiactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC vulnerability, nephrotoxic, antifibrotic and antibacterial activities, and
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.
CC
XX Sequence 464 BP; 112 A; 112 C; 110 G; 126 T; 4 other;
SQ
Query Match 14.6%; Score 18; DB 21; Length 464;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 18 agatgccttaacaacctg 35
|||||
Db 324 agatgccttaacaacctg 341
RESULT 2
AAH29882/C
ID AAH29882 standard; DNA; 1303 BP.
XX
XX AAH29882;
XX
XX 27-JUL-2001 (first entry)
XX
XX C albicans apoptosis associated coding sequence #26.
DE
XX
XX Yeast; fungus; apoptosis; infection; proliferative disease;
KW vaccine; autoimmune disease; ischemia; neurodegeneration; ds.
XX
XX Candida albicans.
OS
XX
XX WO200102550-A2.
PN
XX
XX 11-JAN-2001.
PD
XX
XX 03-JUL-2000; 2000WO-BE00077.
FE
XX
XX 01-JUL-1999; 99EP-0870141.
PR
XX
XX (JANC) JANSEN PHARM NV.
PA
XX
XX Contreras RH, De Backer MD, Luyten WHML, Malcorps IKL;
PI Nelissen BJM, Reekmans RJ;
XX
XX WPI: 2001-367042/38.
DR P-PSDB; AAG70846.
XX
XX Yeast and fungal nucleic acids encoding proteins involved in a pathway
PT leading to programmed cell death, useful for treating proliferative
PT disorders, yeast and fungal infections, or for preventing apoptosis in
PT certain diseases -
XX
PS Claim 17; Fig 2; 218pp; English.

XX The present invention provides the protein and coding sequences of a
CC number of apoptosis associated proteins from the yeast *Saccharomyces*
CC *cerevisiae* and the fungus *Candida albicans*. These can be used to identify
CC treatments for fungal and yeast infections, for proliferative diseases
CC and for apoptosis related diseases such as autoimmune diseases, ischemia
CC and neurodegeneration. The present sequence is one of the *C. albicans*
CC coding sequences of the invention.
XX
SQ Sequence 1303 BP; 488 A; 197 C; 196 G; 422 T; 0 other;
SQ
Query Match 14.6%; Score 18; DB 22; Length 1303;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 106 aattatccacacaaca 123
|||||
Db 138 AATTATTCACACACACA 121
RESULT 3
AAA81461
ID AAA81461 standard; DNA; 9941 BP.
XX
XX AAA81461;
XX
XX 04-DEC-2000 (first entry)
DT
XX
XX N. meningitidis partial DNA sequence gnm_9 SEQ ID NO:9.
DE
XX
XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; genome; immunogenic;
KW antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW *Meningococcus B*; MenB; ds.
XX
XX
XX *Neisseria meningitidis*.
OS
XX
XX WO200022430-A2.
PN
XX
XX 20-APR-2000.
PD
XX
XX 08-OCT-1999; 99WO-US23573.
PF
XX
XX 09-OCT-1998; 98US-0103794.
PR
XX
XX 30-APR-1999; 99US-0132068.
XX
XX (CHIR) CHIRON CORP.
PA
XX
XX Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Massignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;
PI Rappunli R, Pizsa M;
XX
XX WPI: 2000-318079/27.
DR
XX
XX Isolated nucleotide sequences of *Neisseria meningitidis* which can be
PT used in the diagnosis and treatment of *N. meningitidis* infection and
PT other *Neisseria* infections, for example, *N. gonorrhoea* -
XX
PS Claim 7; Page 326-329; 1760pp; English.
XX
XX The present invention describes methods of obtaining immunogenic
CC proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414
CC represent specifically claimed *Neisseria meningitidis* genomic DNA
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to
CC AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF
CC sequences, which are all used in the exemplification of the present
CC invention. The nucleic acid sequences, protein sequences, and antibodies
CC against them, can be used in the manufacture of a composition. The
CC composition can be used as a medicament (or in the manufacture of a
CC medicament) for treating, preventing or diagnosing infection due to

CC Neisserial bacteria. For example, some of the identified proteins could
CC be components of vaccines against *Meningococcus B*; against all serotypes;
CC and/or against all pathogenic *Neisseriae*. Identification of sequences
CC from the bacterium will also facilitate production of biological probes,
CC particularly organism-specific probes. Attempts to make efficacious
CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.
CC Multivalent vaccines have also been tried but none have successfully
CC overcome antigenic variability. The provision of further, complete
CC sequences may provide an opportunity to identify secreted or surface
CC exposed proteins that may be presumed targets for the immune system and
CC which are not antigenically variable or at least more conserved than
CC other more variable regions.
SQ Sequence 9941 BP; 2330 A; 2662 C; 2649 G; 2300 T; 0 other;

Query Match 14.6%; Score 18; DB 21; Length 9941;
Best Local Similarity 100.0%; Pred. No. 2.1;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 97 ctatctccaattatttc 114
Db 7763 ctatctccaattatttc 7780

RESULT 4

AAF21607
ID AAF21607 standard; DNA; 349980 BP.

AC AAF21607;
DT 13-MAR-2001 (first entry)
DE *Neisseria meningitidis B* nucleotide sequence SEQ ID NO:108.

KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; immunogenic; vaccine;
diagnosis; antigen; detection; infection; gene therapy; antibacterial;
ds.

OS *Neisseria meningitidis*.

PN WO200066791-A1.

PD 09-NOV-2000.

PF 08-MAR-2000; 2000MO-US05928.

PR 30-APR-1999; 99US-0132068.

PR 08-OCT-1999; 99WO-US23573.

PR 28-FEB-2000; 2000GB-0004695.

PA (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.

PI Pizze M, Hickey E, Peterson J, Tettelin H, Venter JC, Maignani V;
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V, Rappuoli R;
PI Frazer CM, Grandi G;

XX WPI; 2000-647603/62.

PT *Neisseria meningitidis B* full length genome sequence and open reading
frames are used to detect, treat and prevent *Neisserial* infections -
XX
PS Claim 7; Appendix A; 692pp; English.

CC The present invention describes the full length genome of
CC *Neisseria meningitidis B* (NMB). The sequences in AAF21544 and AAF21607
CC to AAF21613 represent fragments of the NMB genomic sequence, as the
CC sequence was too long to go in a record on its own it was split into 8
CC sequences which overlap each other at the beginning and end of each
CC sequence by 49980 bp (i.e. the last 49980 bp of AAF21544 is repeated at
CC the beginning of AAF21607, the last 49980 bp of AAF21607 are repeated at
CC the beginning of AAF21608, and so on). AAF21545 to AAF21588 encode the

CC *Neisseria* proteins given in AAB58550 to AAB58593, and AAF21589 to
CC AAF21606 represent PCR primers which are used in the exemplification of
CC the present invention. The NMB genome and fragments from it have
CC antibacterial activity, and can be used in vaccines and gene therapy.
CC *Neisseria* nucleic acids, proteins and/or antibodies which binds to the
CC proteins can be used in compositions for treating or preventing infection
CC due to *Neisserial* bacteria or as a diagnostic reagent for detecting the
CC presence of *Neisserial* bacteria or of antibodies raised to *Neisserial*
CC bacteria. Computers, computer memory, computer storage medium or computer
CC databases can be used in a search to identify open reading frames (ORFs)
CC or coding sequences within the NMB genome. The DNA sequences provide
CC further opportunities to find antigenic or immunogenic proteins which are
CC more effective in vaccines than the outer membrane proteins currently
CC used.
SQ Sequence 349980 BP; 84410 A; 84863 C; 94187 G; 86520 T; 0 other;

Query Match 14.6%; Score 18; DB 21; Length 349980;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 97 ctatctccaattatttc 114
Db 103322 ctatctccaattatttc 103339

RESULT 5

AAA81490
ID AAA81490 standard; DNA; 1437668 BP.

AC AAA81490;

DT 04-DEC-2000 (first entry)

DE *N. meningitidis B* full length genome DNA sequence SEQ ID NO:1068.

KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; genome; immunogenic;
antigen; vaccine; diagnosis; infection; antibacterial; identification;

KW *Meningococcus B*; MenB; ds.

OS *Neisseria meningitidis*.

PN WO200022430-A2.

PD 20-APR-2000.

PF 08-OCT-1999; 99WO-US23573.

PR 09-OCT-1998; 98US-0103794.

PR 30-APR-1999; 99US-0132068.

PA (CHIR) CHIRON CORP.

PI Frazer CM, Hickey E, Peterson J, Tettelin H, Venter JC;
PI Maignani V, Galeotti C, Mora M, Ratti G, Scarselli M, Scariato V;
PI Rappuoli R, Pizze M;

XX WPI; 2000-318079/27.

PT Isolated nucleotide sequences of *Neisseria meningitidis* which can be
used in the diagnosis and treatment of *N. meningitidis* infection and
other *Neisserial* infections, for example, *N.gonorrhoea* -
XX
PS Claim 7; Page 866-1272; 1760pp; English.

CC The present invention describes methods of obtaining immunogenic
CC proteins from *Neisseria* genomic sequences. AAA81453 to AAA82414
CC represent specifically claimed *Neisseria meningitidis* genomic DNA
CC sequences; AAA81260 to AAA81303 and AAB25620 to AAB25663 represent
CC *Neisseria* DNA sequences and their corresponding proteins; AAA81254 to
CC AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the
CC isolation of *Neisseria meningitidis* DNA sequences; and AAA81322 to

CC AAA81452 represent *Neisseria meningitidis* MenB polynucleotide ORF
 CC sequences, which are all used in the exemplification of the present
 CC invention. The nucleic acid sequences, protein sequences, and antibodies
 CC against them, can be used in the manufacture of a composition. The
 CC composition can be used as a medicament (or in the manufacture of a
 CC medicament) for treating, preventing or diagnosing infection due to
 CC *Neisseria meningitidis*. For example, some of the identified proteins could
 CC be components of vaccines against *Meningococcus B*; against all serotypes;
 CC and/or against all pathogenic *Neisseria*. Identification of sequences
 CC from the bacterium will also facilitate production of biological probes,
 CC particularly organism-specific probes. Attempts to make efficacious
 CC *Meningococcus B* vaccines have failed mainly due to antigen tolerance.
 CC Multivalent vaccines have also been tried but none have successfully
 CC overcome antigenic variability. The provision of further, complete
 CC sequences may provide an opportunity to identify secreted or surface
 CC exposed proteins that may be presumed targets for the immune system and
 CC which are not antigenically variable or at least more conserved than
 CC other more variable regions.

XX Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other:
 SQ

Query Match 14.6%; Score 18; DB 21; Length 1437668;
 Best Local Similarity 100.0%; Pred. No. 1.9; Mismatches 0; Indels 0; Gaps 0;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 97 ctatctcaattatc 114
 |||||
 Db 403322 ctatctcaattatc 403339

RESULT 6
 AAH29434/C
 ID AAH29434 standard; DNA; 518 BP.
 XX
 XX AAH29434:
 AC
 XX 17-JUL-2001 (first entry)
 DT
 XX
 XX *Drosophila melanogaster* essential gene fragment, SEQ ID NO: 623.
 DE
 XX
 XX *Drosophila melanogaster*; fruit fly; essential gene; screening assay;
 KM
 KW pesticide; crop protection; chromosome 3; ds.
 XX
 OS *Drosophila melanogaster*.
 XX
 PN WO200118547-A1.
 XX
 PD 15-MAR-2001.
 XX
 PF 06-SEP-2000; 2000WO-GB03444.
 XX
 PR 07-SEP-1999; 99GB-0021009.
 XX
 PA (UNIU) UNIV GLASGOW.
 XX
 XX Davies RW, Kaiser K, Yang MY;
 PI
 XX WPI; 2001-281436/29.
 DR
 XX
 PT Screening assays for used for identifying compounds having a
 PT physiological effect on proteins identified as being essential -
 XX
 XX Claim 23; Page 519; 695pp; English.
 PS
 XX
 CC The present sequence is part of an essential gene from *Drosophila*
 CC *melanogaster*. Lack of expression of the protein encoded by this
 CC gene leads to a lethal or semi-lethal phenotype. The invention
 CC relates to 902 nucleic acid sequences from genes encoding proteins
 CC which are thought to be essential, and to a screening assay for
 CC identifying compounds which have a physiological effect on these
 CC proteins. Suitable compounds are useful as pesticides and may be used
 CC in conjunction with other pesticides and herbicides for crop

CC protection. The gene corresponding to the present sequence is located
 CC on chromosome 3.
 CC
 SQ Sequence 518 BP; 134 A; 127 C; 117 G; 140 T; 0 other;

Query Match 13.8%; Score 17; DB 22; Length 518;
 Best Local Similarity 100.0%; Pred. No. 7.6;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 50 caatgcagggcgtgcac 66
 |||||
 Db 208 CAATGCAGGGCTGCAC 192

RESULT 7
 AA103888/C
 ID AA103888 standard; DNA; 446 BP.
 XX
 XX AA103888:
 AC
 XX 09-OCT-2001 (first entry)
 DT
 XX
 XX Probe #879 used to measure gene expression in human breast sample.
 DE
 XX
 XX Probe; human; breast disease; breast cancer; development disorder; ss;
 KW inflammatory disease; proliferative breast disease; non-carcinoma tumour.
 KM
 OS *Homo sapiens*.
 XX
 PN WO200157270-A2.
 XX
 PD 09-AUG-2001.
 XX
 PD 29-JAN-2001; 2001WO-US00661.
 XX
 PF 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 PI
 XX WPI; 2001-476286/51.
 DR
 XX
 PT Novel single exon nucleic acid probe used to measuring gene expression
 PT in a human breast -
 XX
 XX Claim 25; SEQ ID NO 3879; 322pp; English.

CC The present invention relates to novel single exon nucleic acid probes.
 CC The present sequence is one such probe. The probes are useful for
 CC measuring human gene expression in a human breast sample, where the probe
 CC hybridises at high stringency to a nucleic acid expressed in the human
 CC breast. The probes are useful for predicting, diagnosing, grading,
 CC staging, monitoring and prognosing diseases of the human breast,
 CC particularly those diseases with polygenic aetiology. The diseases
 CC include: breast cancer, disorders of development, inflammatory diseases
 CC of the breast, fibrocystic changes, proliferative breast disease and
 CC non-carcinoma tumours.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.

Sequence 446 BP; 126 A; 98 C; 83 G; 139 T; 0 other;

Query Match 13.0%; Score 16; DB 22; Length 446;

Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 74 taagttcacatgtaa 89
|
160 TAAGTTTCACATGTAA 145

RESULT 8

AAI14349/C
ID AAI14349 standard; DNA: 459 BP.

AC AAI14349;

DT 12-OCT-2001 (first entry)

DE Probe #4282 for gene expression analysis in human cervical cell sample.

KW Probe: human; microarray; gene expression; cervical epithelial cell;

KM cervical cancer; ss.

OS Homo sapiens.

PN WO200157278-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00670.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-488901/53.

PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human cervical epithelial cells -

PS Claim 25; SEQ ID No 4282; 487bp; English.

CC The present invention relates to human single exon nucleic acid probes
CC (SENP). The present sequence is one such probe. The SENPs are derived
CC from human HeLa cells. The SENPs can be used to produce a single exon
CC microarray, which can be used for measuring human gene expression in a
CC sample derived from human cervical epithelial cells. By measuring gene
CC expression, the probes are therefore useful in grading and/or staging
CC of diseases of the cervix, notably cervical cancer.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 459 BP; 137 A; 61 C; 110 G; 151 T; 0 other;

Query Match 13.0%; Score 16; DB 22; Length 459;

Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 ccctgatgcactaatg 46
|
409 CCCTGATGACCTAATG 394

RESULT 9

AAI35723/C

ID AAI35723 standard; DNA: 459 BP.

AC AAI35723;

DT 17-OCT-2001 (first entry)

DE Probe #4409 used to measure gene expression in human placenta sample.

KW Probe: microarray; human; placenta; antenatal diagnosis;

KM genetic disorder; ss.

OS Homo sapiens.

PN WO200157272-A2.

PD 09-AUG-2001.

PF 30-JAN-2001; 2001WO-US00663.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

(MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;

DR WPI; 2001-48897/53.

PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -

PS Claim 25; SEQ ID No 4409; 654bp; English.

CC The present invention relates to single exon nucleic acid probes (SENP).
CC The present sequence is one such probe. The probes are useful for
CC producing a microarray for predicting, measuring and displaying gene
CC expression in samples derived from human placenta. The probes are useful
CC for antenatal diagnosis of human genetic disorders.

XX Sequence 459 BP; 137 A; 61 C; 110 G; 151 T; 0 other;

Query Match 13.0%; Score 16; DB 22; Length 459;

Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 31 ccctgatgcactaatg 46
|
409 CCCTGATGACCTAATG 394

RESULT 10

AAI04179/C
ID AAI04179 standard; DNA: 459 BP.

AC AAI04179;

DT 09-OCT-2001 (first entry)

DE Probe #4170 used to measure gene expression in human breast sample.

KW Probe: human; breast disease; breast cancer; development disorder; ss;
KM inflammatory disease; proliferative breast disease; non-carcinoma tumour.

OS Homo sapiens.

PN WO200157270-A2.

PD 09-AUG-2001.
XX
XX 29-JAN-2001; 2001WO-US00061.
XX
XX 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-063366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-476286/51.
XX
XX Novel single exon nucleic acid probe used to measuring gene expression
XX in a human breast -
XX
XX
XX Claim 25; SEQ ID No 4170; 322pp; English.
XX
XX The present invention relates to novel single exon nucleic acid probes.
CC The present sequence is one such probe. The probes are useful for
CC measuring human gene expression in a human breast sample, where the probe
CC hybridizes at high stringency to a nucleic acid expressed in the human
CC breast. The probes are useful for predicting, diagnosing, grading,
CC staging, monitoring and prognosing diseases of the human breast,
CC particularly those diseases with polygenic aetiology. The diseases
CC include: breast cancer, disorders of development, inflammatory diseases
CC of the breast, fibrocystic changes, proliferative breast disease and
CC non-carcinoma tumours.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 459 BP; 137 A; 61 C; 110 G; 151 T; 0 other;
SO

Query Match 13.0%; Score 16; DB 22; Length 459;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 31 cccgtatgacctaatg 46
|||||
DB 409 CCTGATGACTAATG 394

RESULT 11
AA34755/c
ID AA34755 standard; DNA; 852 BP.
XX
XX AA34755;
XX
XX 05-JUL-1999 (first entry)
DT
XX
XX DNA encoding OMP-1W protein.
DE
XX
XX Outer membrane protein; OMP; *Escherichia chaffeensis*; E. canis; P30;
KW detection; dog; ss.
XX
XX *Escherichia chaffeensis*.
OS
XX
XX WO9913720-A1.
PN
XX
XX 25-MAR-1999.
PD
XX
XX 18-SEP-1998; 98WO-US19600.
PF
XX
XX 19-SEP-1997; 97US-0059353.
PR
XX
XX (OHIS) UNIV OHIO STATE.
PA

XX
XX Ohashi N, Rikihisa Y;
PI
XX
XX WPI; 1999-254290/21.
DR
XX
XX P-PSDB; AAY06955.
DR
XX
XX Novel outer membrane proteins from *Escherichia chaffeensis* and
PT *Escherichia canis*
XX
XX
XX Disclosure; Fig 15A; 55pp; English.
PS
XX
XX The invention provides isolated outer membrane proteins (OMP) from
CC *Escherichia chaffeensis* and *E. canis*. The E. chaffeensis proteins form part
CC of the OMP family and consist of proteins OMP-1, -1(B to Z) shown
CC in AAY06943-958. The E. canis proteins form part of the P30 family and
CC consist of proteins shown in AAY06959-970. The proteins and genes are
CC used to detect E. chaffeensis in patients and E. canis in dogs.
XX
XX
XX Sequence 852 BP; 295 A; 115 C; 141 G; 301 T; 0 other;
SO

Query Match 13.0%; Score 16; DB 20; Length 852;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 84 atgaagtgtatctt 99
|||||
DB 578 ATGTAAGTGTAATCTT 563

RESULT 12
AA38323/c
ID AA38323 standard; DNA; 1278 BP.
XX
XX AA38323;
XX
XX 21-AUG-2000 (first entry)
DT
XX
XX Human angiotensinogen (AGT) gene regulatory sequence (GenBank X15323).
DE
XX
XX Angiotensinogen gene; AGT; regulatory region; polymorphism;
KW polymorphic marker; cardiovascular disease; myocardial infarction;
KW unstable angina; hypertension; atherosclerosis; stroke; prognosis;
KW drug screening; treatment outcome; human; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200022166-A2.
PN
XX
XX 20-APR-2000.
PD
XX
XX 13-OCT-1999; 99WO-IB01678.
PF
XX
XX 14-OCT-1998; 98US-0104286.
PR
XX
XX 14-OCT-1998; 98US-0104302.
XX
XX (EURO-) EUROPA MEDICAL AB.
PA
XX
XX Norberg LT, Andersson MK, Lindstrom PHR, Jonsson L;
PI
XX
XX WPI; 2000-318010/27.
DR
XX
XX
XX Assessing cardiovascular status in humans involves comparing test
PT polymorphic pattern comprising polymorphic positions within genes
PT encoding specific proteins, with reference polymorphic pattern -
XX
XX
XX Disclosure; Page -; 126pp; English.
PS
XX
XX The invention relates to a novel method of assessing the cardiovascular
CC status in an individual and to newly identified polymorphisms in the
CC genes encoding angiotensin-converting enzyme (ACE), angiotensin II
CC receptor type 1 (AT1) and type 2 (AT2), angiotensinogen (AGT), renin,
CC aldosterone synthase, endothelin receptor type A and beta-adrenergic

Query Match	13.0%	Score 16	DB 21	Length 1278
Best Local Similarity	100.0%	Pred. NO. 26		
Matches 16	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	71	tggttaagtttcacatg	86	
Db	522	TGTTAAGTTTCACATG	507	
RESULT 13				
AAA8324/c				
ID	AAA8324	standard; DNA; 1278 BP.		
XX				
AC	AAA8324;			
XX				
DT	21-AUG-2000	(first entry)		
XX				
DE				
XX	Human angiotensinogen (AGT) gene exon 2.			
XX				
KW	Angiotensinogen gene; AGT; coding region; polymorphism;			
KW	polymorphic marker; cardiovascular disease; myocardial infarction;			
KW	unstable angina; hypertension; atherosclerosis; stroke; prognosis;			
KW	drug screening; treatment outcome; human; ds.			
XX				
OS	Homo sapiens.			
XX				
PN	W0200022166-A2.			
XX				
PD	20-APR-2000.			
XX				
PF	13-OCT-1999;	99WO-IB01678.		
XX				
14-OCT-1998;	98US-0104286.			
XX				

The invention relates to a novel method of assessing the cardiovascular status in an individual and to newly identified polymorphisms in the genes encoding angiotensin-converting enzyme (ACE), angiotensin II receptor type 1 (AT1) and type 2 (AT2), angiotensinogen (AGT), renin, aldoesterone synthase, endothelin receptor type A and beta-adrenergic receptors 1 and 2. The method comprises determining the sequence at one or more polymorphic positions within these genes, and comparing the pattern of polymorphisms from the individual with a reference polymorphic pattern obtained from a population of individuals exhibiting a predetermined cardiovascular disease status. The polymorphic markers are useful for determining the predisposition of an individual to cardiovascular disorders such as myocardial infarction, unstable angina, hypertension, atherosclerosis and stroke. They are also useful for predicting the likely cardiovascular status of a patient given a treatment regimen comprising administration of cardiovascular drugs (e.g., ACE inhibitors, beta-adrenergic receptor antagonists (beta-blockers) or calcium channel blockers). One or more polymorphic markers provides a basis for predicting the outcome of a treatment regimen. Fragments of the genes comprising a polymorphic site may be used as primers and probes for detecting genetic polymorphisms or in molecular library arrays for high throughput screening. The genes, and the proteins they encode are useful in the screening of potential cardiovascular drugs. Determination of an individual's polymorphic pattern reduces or eliminates trial and error in selecting a treatment for a particular individual cardiovascular patient. It also provides the ability to eliminate patients from clinical trials who are predicted to be non-responsive, or at a risk for an adverse response, to a particular treatment regimen. Adverse results in an early trial can be evaluated to identify polymorphic patterns so that the adverse results can be correlated with a sub-population of the test population, permitting exclusion of such sub-populations from the treatment group. Beneficial drugs can be approved for use in the appropriate population, thereby decreasing the number of patients required for a clinical trial, which in turn decreases the duration and cost of such trials. Sequences A38324-A38327 represent exons 2, 3, 4 and 5 of the human angiotensinogen gene coding region (GenBank M24686-M24689). The polymorphic sites identified are 273C/T, 620C/T, 803T/C, 912C/T, 997G/C, 11116G/A, 1174C/A; and A or G at position 49 in exon 4.

Sequence 1278 BP; 283 A; 343 C; 295 G; 357 T; 0 other;

Query Match	13.0%;	Score 16;	DB 21;	Length 1278;
Best Local Similarity	100.0%;	Pred. No. 26;		
Matches 16;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
OY	71	tggttaagtttcacatg	86	
Db	522	TGGTAAGTTTCACATG	507	
RESULT 14				
AAQ47075				
ID	AAQ47075	standard;	DNA; 1491	BP.
XX				
AC	AAQ47075;			
XX				
DT	13-JAN-1994	(first entry)		
XX				

DE 21B4/rhoptry antigen gene 5 DNA.
 XX
 XX Polymerase chain reaction; PCR; amplify; primer; detection;
 KM babesiosis; parasite; Babesia bovis; 21B4/rhoptry; antigen; gene;
 KM repeat region; immune response; vaccine; ss.
 XX
 OS Synthetic.
 XX
 XX MO9314204-A.
 XX
 XX 22-JUL-1993.
 XX
 XX 15-JAN-1993; 93MO-AU00012.
 XX
 XX 15-JAN-1992; 92AU-0000399.
 XX
 XX (CSIR) COMMONWEALTH SCI & IND RES ORG.
 PA
 PI Dalrymple BP, Peters JW;
 XX
 XX WPI: 1993-243219/30.
 DR
 DR P-PSDB; AAR39901.
 XX
 XX
 PT Detecting closely linked gene copies which encode protective
 PT antigen against babesiosis - by screening babesial genomic DNA
 PT library with oligo-nucleotide probe based partial sequencing of
 PT protective antigen and identifying positive clones
 PS
 PS Claim 23; Fig 6; 55pp; English.
 XX
 XX This sequence represents the Babesia bovis 21B4/rhoptry antigen gene
 CC 5. This sequence was isolated by PCR using the primers given in
 CC AA047068-72. Primer 21B4.1 corresponds to part of the repeated
 CC region of 21B4/rhoptry antigen. In hybridisation assays this primer
 CC recognised two tandemly repeated regions suggesting that B. bovis
 CC contains two copies of the 21B4/rhoptry antigen gene. The two
 CC proteins encoded by the two antigen genes are identical. Primers
 CC 21B4.2 and 21B4.3 flank the 21B4-309 coding region of the antigen
 CC gene. Primer 21B4.4 primes synthesis just 3' to the end of the open
 CC reading frame. The entire open reading frame was shown to encode five
 CC antigen genes. The 3' non-repetitive sequences of open reading frames
 CC 1-4 are identical. Gene 5 shows sequence divergence throughout most
 CC of the open reading frame. Babesia antigen genes can be used in the
 CC production of a combined vaccine which will stimulate a greater immune
 CC response and afford broader immunity than a single antigen vaccine.
 CC
 XX
 XX Sequence 1491 BP; 479 A; 286 C; 328 G; 398 T; 0 other;
 SQ
 Query Match 13.0%; Score 16; DB 14; Length 1491;
 Best Local Similarity 100.0%; Pred. No. 26;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 24 cttaacaccctcatga 39
 Db 941 cttaacaccctcatga 956
 ||||||||||||||||
 RESULT 15
 AAC98971/c
 ID AAC98971 standard; CDNA: 2732 BP.
 XX
 AC AAC98971;
 XX
 DT 09-MAR-2001 (first entry)
 XX
 DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:199.
 XX
 XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;
 KM detection; diagnosis; identification; cytosolic; neuroprotective;
 KM neotrophic; immunomodulatory; relaxant; contraceptive; gynaecological;
 KM antiinflammatory; cardiac; gene therapy; chromosome mapping;
 KM linkage analysis; tissue identification; tissue typing; forensic;

KM neural; immune system; muscular; reproductive; gastrointestinal;
 KM pulmonary; cardiovascular; renal; proliferative; ss.
 XX
 XX Homo sapiens.
 OS
 XX
 XX MO200055320-A1.
 XX
 XX 21-SEP-2000.
 XX
 XX 08-MAR-2000; 2000MO-US05989.
 XX
 XX 12-MAR-1999; 99US-0124270.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA
 PI Rosen CA, Ruben SM;
 XX
 XX WPI: 2000-579444/54.
 DR
 DR P-PSDB; AAB54206.
 XX
 XX
 PT New nucleic acid that is a pancreatic cancer antigen for preventing,
 PT treating, or ameliorating a medical condition, particular pancreatic
 PT cancer, or for use in assays for diagnosing a pathological condition -
 XX
 XX
 PS Claim 1; Page 645-646; 1379pp; English.
 XX
 XX AAC98773 to AAC99231 encode the human pancreatic cancer associated
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to
 CC AAB54466. The human pancreatic cancer antigens have cytosolic,
 CC neuroprotective, neotrophic, immunomodulatory, relaxant, contraceptive,
 CC gynaecological, cardiac and antiinflammatory activities, and can be used
 CC in gene therapy. The polynucleotide and proteins can be used for
 CC preventing, treating, or ameliorating a medical condition or in assays
 CC for diagnosing a pathological condition or a susceptibility to one in a
 CC subject. Binding partners to the proteins and the activity of the
 CC proteins can be identified. The pancreatic cancer antigens can be used to
 CC detect, treat or prevent pancreatic disorders, especially cancer.
 CC Agonists and antagonists to the antigens can be screened for. The
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic
 CC acid hybridisation probes that can be used in chromosome mapping, linkage
 CC analysis, tissue identification and/or typing and a variety of forensic
 CC and diagnostic methods. The proteins can be used to generate antibodies
 CC which are used to purify, detect and target the polypeptides, including
 CC both in vivo and in vitro diagnostic and therapeutic methods. The
 CC proteins can be used to treat or prevent neural, immune system, muscular,
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent
 CC sequences used in the exemplification of the present invention.
 CC
 XX
 XX Sequence 2732 BP; 624 A; 753 C; 688 G; 659 T; 8 other;
 SQ
 Query Match 13.0%; Score 16; DB 21; Length 2732;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 101 tctcaattattcca 116
 Db 1444 tctcaattattcca 1429
 ||||||||||||||||
 RESULT 16
 AAC78084/c
 ID AAC78084 standard; CDNA: 4201 BP.
 XX
 AC AAC78084;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human cancer associated gene sequence SEQ ID NO:478.
 XX
 XX Human; cancer associated gene; cancer antigen; detection; cancer;
 KM diagnosis; cytosolic; proliferative; vulnerrary; immunomodulator;

KW antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW antiinflammatory; antihypertoid; antiatherogenic; antibacterial; cardiant;
KW dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW immune disorder; hematopoietic cell disorder; autoimmune disorder;
KW allergic reaction; graft versus host disease; organ rejection;
KW haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening; ss.
XX
OS Homo sapiens.
XX
PN WO20005350-A1.
XX
PD 21-SEP-2000.
XX
PF 08-MAR-2000; 2000WO-US05882.
XX
PR 12-MAR-1999; 99US-0124270.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Ruben SM;
XX
DR WPI: 2000-587533/55.
DR P-PSDB: AAB43875.
XX
PT Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -
XX
PS Claim 1; Page 1012-1013; 2352pp; English.
XX
CC AAC77607 to AAC78448 encode the human cancer associated proteins given
CC in AAB43398 to AAB44239. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnerary; immunomodulator;
CC antidiabetic; antiasthmatic; antirheumatic; antiarthritic;
CC antiinflammatory; antihypertoid; antiatherogenic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC nootropic; vasotropic; antipsoriatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of hematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies,
CC agonists and antagonists may be also be used in drug screens. AAC78449 to
CC AAC78457 and AAB44240 represent sequences used in the exemplification of
CC the present invention.
XX
SQ Sequence 4201 BP; 1323 A; 637 C; 850 G; 1381 T; 10 other;

Query Match 13.0%; Score 16; DB 21; Length 4201;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 aggtatgtaaatg 17
|||||
Db 2555 AGGTATGTGTAATGTG 2540

RESULT 17
AAH3277/c
ID AAH33277 standard; cDNA; 4201 BP.
XX
AC AAH33277;
XX
DT 03-SEP-2001 (first entry)
XX

DE Human colon cancer antigen encoding cDNA SEQ ID NO:333.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX
XX Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI: 2001-235357/24.
DR P-PSDB: AAG73846.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 1; Page 2445-2447; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 4201 BP; 1323 A; 637 C; 850 G; 1381 T; 10 other;

Query Match 13.0%; Score 16; DB 22; Length 4201;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 aggtatgtaaatg 17
|||||
Db 2555 AGGTATGTGTAATGTG 2540

RESULT 18
AAC91600/c
ID AAC91600 standard; DNA; 5308 BP.
XX
AC AAC91600;
XX
DT 16-MAR-2001 (first entry)
XX
DE Human angiotensinogen gene 5' region/exon 1, SEQ ID NO:1.
XX
KW Human angiotensinogen gene; AGT; insulin-dependent diabetes mellitus;
KW type 1 diabetes; chromosome 1q42-43; single nucleotide polymorphism;
KW IDDM; SNP; diagnosis; susceptibility; transgenic animal; drug screening;

AAC91671/c
 ID AAC91671 standard; DNA: 5308 BP.
 XX
 AC AAC91671:
 XX
 DT 16-MAR-2001 (first entry)
 XX
 DE Human angiotensinogen gene 5' region mutant, T2271C.
 XX
 KW Human angiotensinogen gene; AGT; insulin-dependent diabetes mellitus;
 KW type 1 diabetes; chromosome 1q42-43; single nucleotide polymorphism;
 KW IDDM; SNP; diagnosis; susceptibility; transgenic animal; drug screening;
 KW antidiabetic; gene therapy; 5' region; mutant; exon 1; ds.
 XX
 OS Homo sapiens.
 OS
 PN WO200071751-A1.
 XX
 PD 30-NOV-2000.
 XX
 PF 16-MAY-2000; 2000WO-US13327.
 XX
 PR 21-MAY-1999; 99US-0135423.
 PR 06-JAN-2000; 2000US-0174700.
 XX
 PA (MYRI-) MYRIAD GENETICS INC.
 PI McGrail M, Russell DL, Shattuck DM;
 DR WPI; 2001-025172/03.
 XX
 PT Novel angiotensinogen gene, mutant alleles of which causes
 PT susceptibility to insulin-dependent diabetes mellitus useful for
 PT diagnosis of predisposition to diabetes -
 PS
 PS Claim 2; Page -: 83pp; English.
 XX
 CC The invention relates to the human angiotensinogen (AGT) gene, some
 CC mutant alleles of which cause a susceptibility to insulin-dependent
 CC diabetes mellitus (IDDM, type 1 diabetes). The AGT gene is located
 CC on chromosome 1q42-43, a region linked to IDDM. The invention
 CC discloses genomic sequences comprising exons 1-5 of the human AGT gene
 CC (AAC91600-C91604) and a genomic sequence comprising an alternative AGT
 CC gene exon 1 (AAC91606). The invention also encompasses the specifically
 CC claimed human AGT mutant nucleic acid sequences AAC91667-C91684, and the
 CC mutant angiotensinogen proteins ABA8945-B48949. The invention also
 CC relates to detecting mutant AGT alleles or gene products thereof which
 CC are related to IDDM; determining whether a person has, or is at risk of
 CC developing diabetes via detection of a polymorphism in the AGT gene; and
 CC methods of screening for drug candidates which may be useful in the
 CC treatment of diabetes resulting from an AGT mutation. Methods of
 CC preventing or treating diabetes are claimed which comprise the
 CC administration of a compound which agonises or antagonises wild-type or
 CC mutant AGT, which agonises or antagonises an AGT receptor, which
 CC inhibits AGT gene expression, or which cleaves AGT proteins. In addition,
 CC the invention encompasses a transgenic non-human animal, or cell line
 CC derived therefrom, comprising a mutant human AGT allele. The
 CC polymorphisms identified in the AGT gene are useful for determining if a
 CC person has, or is at risk from developing insulin-dependent diabetes
 CC mellitus. AGT modulators can be used to treat or prevent diabetes. Mutant
 CC AGT proteins or fragments thereof are useful for screening compounds
 CC which bind to AGT polypeptides. The present sequence represents a portion
 CC of the human AGT gene comprising a mutant 5' region fragment, and exon
 CC 1. Note: The present sequence is not shown in the specification, but is
 CC derived from the wild-type human AGT gene 5' region/exon 1 sequence
 CC shown on page 60-63.
 XX
 SO Sequence 5308 BP; 1364 A; 1266 C; 1329 G; 1349 T; 0 other;

Query Match 13.0%; Score 16; DB 22; Length 5308;
 Best Local Similarity 100.0%; Pred. No. 25;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 tgtraagttcacatg 86
 ||||||||||||||||
 Db 2311 TGTRAAGTTTCACATG 2296
 RESULT 21
 AAC91672/c
 ID AAC91672 standard; DNA: 5308 BP.
 XX
 AC AAC91672:
 XX
 DT 16-MAR-2001 (first entry)
 XX
 DE Human angiotensinogen gene 5' region mutant, C2731T.
 XX
 KW Human angiotensinogen gene; AGT; insulin-dependent diabetes mellitus;
 KW type 1 diabetes; chromosome 1q42-43; single nucleotide polymorphism;
 KW IDDM; SNP; diagnosis; susceptibility; transgenic animal; drug screening;
 KW antidiabetic; gene therapy; 5' region; mutant; exon 1; ds.
 XX
 OS Homo sapiens.
 OS
 PN WO200071751-A1.
 XX
 PD 30-NOV-2000.
 XX
 PF 16-MAY-2000; 2000WO-US13327.
 XX
 PR 21-MAY-1999; 99US-0135423.
 PR 06-JAN-2000; 2000US-0174700.
 XX
 PA (MYRI-) MYRIAD GENETICS INC.
 PI McGrail M, Russell DL, Shattuck DM;
 DR WPI; 2001-025172/03.
 XX
 PT Novel angiotensinogen gene, mutant alleles of which causes
 PT susceptibility to insulin-dependent diabetes mellitus useful for
 PT diagnosis of predisposition to diabetes -
 PS
 PS Claim 2; Page -: 83pp; English.
 XX
 CC The invention relates to the human angiotensinogen (AGT) gene, some
 CC mutant alleles of which cause a susceptibility to insulin-dependent
 CC diabetes mellitus (IDDM, type 1 diabetes). The AGT gene is located
 CC on chromosome 1q42-43, a region linked to IDDM. The invention
 CC discloses genomic sequences comprising exons 1-5 of the human AGT gene
 CC (AAC91600-C91604) and a genomic sequence comprising an alternative AGT
 CC gene exon 1 (AAC91606). The invention also encompasses the specifically
 CC claimed human AGT mutant nucleic acid sequences AAC91667-C91684, and the
 CC mutant angiotensinogen proteins ABA8945-B48949. The invention also
 CC relates to detecting mutant AGT alleles or gene products thereof which
 CC are related to IDDM; determining whether a person has, or is at risk of
 CC developing diabetes via detection of a polymorphism in the AGT gene; and
 CC methods of screening for drug candidates which may be useful in the
 CC treatment of diabetes resulting from an AGT mutation. Methods of
 CC preventing or treating diabetes are claimed which comprise the
 CC administration of a compound which agonises or antagonises wild-type or
 CC mutant AGT, which agonises or antagonises an AGT receptor, which
 CC inhibits AGT gene expression, or which cleaves AGT proteins. In addition,
 CC the invention encompasses a transgenic non-human animal, or cell line
 CC derived therefrom, comprising a mutant human AGT allele. The
 CC polymorphisms identified in the AGT gene are useful for determining if a
 CC person has, or is at risk from developing insulin-dependent diabetes
 CC mellitus. AGT modulators can be used to treat or prevent diabetes. Mutant
 CC AGT proteins or fragments thereof are useful for screening compounds
 CC which bind to AGT polypeptides. The present sequence represents a portion
 CC of the human AGT gene comprising a mutant 5' region fragment, and exon
 CC 1. Note: The present sequence is not shown in the specification, but is
 CC derived from the wild-type human AGT gene 5' region/exon 1 sequence
 CC shown on page 60-63.

XX Sequence 5308 BP; 1364 A; 1264 C; 1329 G; 1351 T; 0 other;
SQ
Query Match 13.0%; Score 16; DB 22; Length 5308;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 71 tggtaagttcacatg 86
|||||
Db 2311 TGGTAAGTTCAACATG 2296

RESULT 22
AAC91674/C
ID AAC91674 standard; DNA; 5308 BP.
XX
AC AAC91674;
XX
DT 16-MAR-2001 (first entry)
XX
DE Human angiotensinogen gene 5' region mutant, T2649A.
XX
KW Human angiotensinogen gene; AGT; insulin-dependent diabetes mellitus;
KW type 1 diabetes; chromosome 1q42-43; single nucleotide polymorphism;
KW IDDM; SNP; diagnosis; susceptibility; transgenic animal; drug screening;
KW antidiabetic; gene therapy; 5' region; mutant; exon 1; ds.
XX
OS Homo sapiens.
XX
PN WO200071751-A1.
XX
PD 30-NOV-2000.
XX
PF 16-MAY-2000; 2000WO-US13327.
XX
PR 21-MAY-1999; 990US-0135423.
PR 06-JAN-2000; 2000US-0174700.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI McGrail M, Russell DL, Shattuck DM;
XX
DR WPI; 2001-025172/03.
XX
PT Novel angiotensinogen gene, mutant alleles of which causes
PT susceptibility to insulin-dependent diabetes mellitus useful for
PT diagnosis of predisposition to diabetes -
XX
XX
XX Claim 2; Page -: 83pp; English.
XX
XX The invention relates to the human angiotensinogen (AGT) gene, some
XX mutant alleles of which cause a susceptibility to insulin-dependent
XX diabetes mellitus (IDDM, type 1 diabetes). The AGT gene is located
XX on chromosome 1q42-43, a region linked to IDDM. The invention
XX discloses genomic sequences comprising exons 1-5 of the human AGT gene
XX (AAC91600-C91604) and a genomic sequence comprising an alternative AGT
XX gene exon 1 (AAC91606). The invention also encompasses the specifically
XX claimed human AGT mutant nucleic acid sequences AAC91667-C91684, and the
XX mutant angiotensinogen proteins AAB48945-B48949. The invention also
XX relates to detecting mutant AGT alleles or gene products thereof which
XX are related to IDDM; determining whether a person has, or is at risk of
XX developing diabetes via detection of a polymorphism in the AGT gene; and
XX methods of screening for drug candidates which may be useful in the
XX treatment of diabetes resulting from an AGT mutation. Methods of
XX preventing or treating diabetes are claimed which comprise the
XX administration of a compound which agonises or antagonises wild-type or
XX mutant AGT, which agonises or antagonises an AGT receptor, which
XX inhibits AGT gene expression, or which cleaves AGT proteins. In addition,
XX the invention encompasses a transgenic non-human animal, or cell line
XX derived therefrom, comprising a mutant human AGT allele. The
XX polymorphisms identified in the AGT gene are useful for determining if a
XX person has, or is at risk from developing insulin-dependent diabetes

CC mellitus. AGT modulators can be used to treat or prevent diabetes. Mutant
CC AGT proteins or fragments thereof are useful for screening compounds
CC which bind to AGT polypeptides. The present sequence represents a portion
CC of the human AGT gene comprising a mutant 5' region fragment, and exon
CC 1. Note: The present sequence is not shown in the specification, but is
CC derived from the wild-type human AGT gene 5' region/exon 1 sequence
CC shown on page 60-63.
XX
SQ Sequence 5308 BP; 1365 A; 1265 C; 1329 G; 1349 T; 0 other;
OY 71 tggtaagttcacatg 86
|||||
Db 2311 TGGTAAGTTCAACATG 2296

RESULT 23
AAC91675/C
ID AAC91675 standard; DNA; 5308 BP.
XX
AC AAC91675;
XX
DT 16-MAR-2001 (first entry)
XX
DE Human angiotensinogen gene 5' region mutant, G2753A.
XX
KW Human angiotensinogen gene; AGT; insulin-dependent diabetes mellitus;
KW type 1 diabetes; chromosome 1q42-43; single nucleotide polymorphism;
KW IDDM; SNP; diagnosis; susceptibility; transgenic animal; drug screening;
KW antidiabetic; gene therapy; 5' region; mutant; exon 1; ds.
XX
OS Homo sapiens.
XX
PN WO200071751-A1.
XX
PD 30-NOV-2000.
XX
PF 16-MAY-2000; 2000WO-US13327.
XX
PR 21-MAY-1999; 990US-0135423.
PR 06-JAN-2000; 2000US-0174700.
XX
PA (MYRI-) MYRIAD GENETICS INC.
XX
PI McGrail M, Russell DL, Shattuck DM;
XX
DR WPI; 2001-025172/03.
XX
PT Novel angiotensinogen gene, mutant alleles of which causes
PT susceptibility to insulin-dependent diabetes mellitus useful for
PT diagnosis of predisposition to diabetes -
XX
XX
XX Claim 2; Page -: 83pp; English.
XX
XX The invention relates to the human angiotensinogen (AGT) gene, some
XX mutant alleles of which cause a susceptibility to insulin-dependent
XX diabetes mellitus (IDDM, type 1 diabetes). The AGT gene is located
XX on chromosome 1q42-43, a region linked to IDDM. The invention
XX discloses genomic sequences comprising exons 1-5 of the human AGT gene
XX (AAC91600-C91604) and a genomic sequence comprising an alternative AGT
XX gene exon 1 (AAC91606). The invention also encompasses the specifically
XX claimed human AGT mutant nucleic acid sequences AAC91667-C91684, and the
XX mutant angiotensinogen proteins AAB48945-B48949. The invention also
XX relates to detecting mutant AGT alleles or gene products thereof which
XX are related to IDDM; determining whether a person has, or is at risk of
XX developing diabetes via detection of a polymorphism in the AGT gene; and
XX methods of screening for drug candidates which may be useful in the
XX treatment of diabetes resulting from an AGT mutation. Methods of
XX preventing or treating diabetes are claimed which comprise the

CC administration of a compound which agonises or antagonises wild-type or
CC mutant AGT, which agonises or antagonises an AGT receptor, which
CC inhibits AGT gene expression, or which cleaves AGT proteins. In addition,
CC the invention encompasses a transgenic non-human animal, or cell line
CC derived therefrom, comprising a mutant human AGT allele. The
CC polymorphisms identified in the AGT gene are useful for determining if a
CC person has, or is at risk from developing insulin-dependent diabetes
CC mellitus. AGT modulators can be used to treat or prevent diabetes. Mutant
CC AGT proteins or fragments thereof are useful for screening compounds
CC which bind to AGT polypeptides. The present sequence represents a portion
CC of the human AGT gene comprising a mutant 5' region fragment, and exon
CC 1. Note: The present sequence is not shown in the specification, but is
CC derived from the wild-type human AGT gene 5' region/exon 1 sequence
CC shown on page 60-63.

SO Sequence 5308 BP; 1365 A; 1265 C; 1328 G; 1350 T; 0 other;

Query Match 13.0%; Score 16; DB 22; Length 5308;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 71 tggtaagttcacatg 86
|||||
Db 2311 TGCTAAGTTTCACATG 2296

RESULT 24
AAC91676/c
ID AAC91676 standard; DNA; 5308 BP.
XX AAC91676;
XX
DT 16-MAR-2001 (first entry)
XX
XX Human angiotensinogen gene 5' region mutant, T2759C.
DE
XX Human angiotensinogen gene; AGT; insulin-dependent diabetes mellitus;
KW type 1 diabetes; chromosome 1q42-43; single nucleotide polymorphism;
KM IDDM; SNP; diagnosis; susceptibility; transgenic animal; drug screening;
XX antidiabetic; gene therapy; 5' region; mutant; exon 1; ds.
XX
XX Homo sapiens.
OS
PN WO200071751-A1.
XX
PD 30-NOV-2000.
XX
PF 16-MAY-2000; 2000WO-US13327.
XX
PR 21-MAY-1999; 99US-0135423.
PR 06-JAN-2000; 2000US-0174700.
XX
XX (MYRI-) MYRIAD GENETICS INC.
XX
XX McGrail M, Russell DL, Shattuck DM;
PI
XX WPI; 2001-025172/03.
DR
XX
PT Novel angiotensinogen gene, mutant alleles of which causes
PT susceptibility to insulin-dependent diabetes mellitus useful for
PT diagnosis of predisposition to diabetes -
XX
XX Claim 2; Page -: 83pp; English.

CC The invention relates to the human angiotensinogen (AGT) gene, some
CC mutant alleles of which cause a susceptibility to insulin-dependent
CC diabetes mellitus (IDDM, type 1 diabetes). The AGT gene is located
CC on chromosome 1q42-43, a region linked to IDDM. The invention
CC discloses genomic sequences comprising exons 1-5 of the human AGT gene
CC (AAC91600-C91604) and a genomic sequence comprising an alternative AGT
CC gene exon 1 (AAC91606). The invention also encompasses the specifically
CC claimed human AGT mutant nucleic acid sequences AAC91667-C91684, and the

CC mutant angiotensinogen proteins AB48945-B48949. The invention also
CC relates to detecting mutant AGT alleles or gene products thereof which
CC are related to IDDM; determining whether a person has, or is at risk of
CC developing diabetes via detection of a polymorphism in the AGT gene; and
CC methods of screening for drug candidates which may be useful in the
CC treatment of diabetes resulting from an AGT mutation. Methods of
CC preventing or treating diabetes are claimed which comprise the
CC administration of a compound which agonises or antagonises wild-type or
CC mutant AGT, which agonises or antagonises an AGT receptor, which
CC inhibits AGT gene expression, or which cleaves AGT proteins. In addition,
CC the invention encompasses a transgenic non-human animal, or cell line
CC derived therefrom, comprising a mutant human AGT allele. The
CC polymorphisms identified in the AGT gene are useful for determining if a
CC person has, or is at risk from developing insulin-dependent diabetes
CC mellitus. AGT modulators can be used to treat or prevent diabetes. Mutant
CC AGT proteins or fragments thereof are useful for screening compounds
CC which bind to AGT polypeptides. The present sequence represents a portion
CC of the human AGT gene comprising a mutant 5' region fragment, and exon
CC 1. Note: The present sequence is not shown in the specification, but is
CC derived from the wild-type human AGT gene 5' region/exon 1 sequence
CC shown on page 60-63.

SO Sequence 5308 BP; 1364 A; 1266 C; 1329 G; 1349 T; 0 other;

Query Match 13.0%; Score 16; DB 22; Length 5308;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 71 tggtaagttcacatg 86
|||||
Db 2311 TGCTAAGTTTCACATG 2296

RESULT 25
AAC91677/c
ID AAC91677 standard; DNA; 5308 BP.
XX AAC91677;
XX
DT 16-MAR-2001 (first entry)
XX
XX Human angiotensinogen gene 5' region mutant, G2829A.
DE
XX Human angiotensinogen gene; AGT; insulin-dependent diabetes mellitus;
KW type 1 diabetes; chromosome 1q42-43; single nucleotide polymorphism;
KM IDDM; SNP; diagnosis; susceptibility; transgenic animal; drug screening;
XX antidiabetic; gene therapy; 5' region; mutant; exon 1; ds.
XX
XX Homo sapiens.
OS
PN WO200071751-A1.
XX
PD 30-NOV-2000.
XX
PF 16-MAY-2000; 2000WO-US13327.
XX
PR 21-MAY-1999; 99US-0135423.
PR 06-JAN-2000; 2000US-0174700.
XX
XX (MYRI-) MYRIAD GENETICS INC.
XX
XX McGrail M, Russell DL, Shattuck DM;
PI
XX WPI; 2001-025172/03.
DR
XX
PT Novel angiotensinogen gene, mutant alleles of which causes
PT susceptibility to insulin-dependent diabetes mellitus useful for
PT diagnosis of predisposition to diabetes -
XX
XX Claim 2; Page -: 83pp; English.

CC The invention relates to the human angiotensinogen (AGT) gene, some

CC The invention relates to the human angiotensinogen (AGT) gene, some
 CC mutant alleles of which cause a susceptibility to insulin-dependent
 CC diabetes mellitus (IDDM, type 1 diabetes). The AGT gene is located
 CC on chromosome 1q42-43, a region linked to IDDM. The invention
 CC discloses genomic sequences comprising exons 1-5 of the human AGT gene
 CC (AAC91600-C91604) and a genomic sequence comprising an alternative AGT
 CC gene exon 1 (AAC91605). The invention also encompasses the specifically
 CC claimed human AGT mutant nucleic acid sequences AAC91667-C91684, and the
 CC mutant angiotensinogen proteins AAB48945-B48949. The invention also
 CC relates to detecting mutant AGT alleles or gene products thereof which
 CC are related to IDDM, determining whether a person has, or is at risk of
 CC developing diabetes via detection of a polymorphism in the AGT gene; and
 CC methods of screening for drug candidates which may be useful in the
 CC treatment of diabetes resulting from an AGT mutation. Methods of
 CC preventing or treating diabetes are claimed which comprise the
 CC administration of a compound which agonises or antagonises wild-type or
 CC mutant AGT, which agonises or antagonises an AGT receptor, which
 CC inhibits AGT gene expression, or which cleaves AGT proteins. In addition,
 CC the invention encompasses a transgenic non-human animal, or cell line
 CC derived therefrom, comprising a mutant human AGT allele. The
 CC polymorphisms identified in the AGT gene are useful for determining if a
 CC person has, or is at risk from developing insulin-dependent diabetes
 CC mellitus. AGT modulators can be used to treat or prevent diabetes. Mutant
 CC AGT proteins or fragments thereof are useful for screening compounds
 CC which bind to AGT polypeptides. The present sequence represents a human
 CC AGT gene 5' region/promoter PCR primer used in an exemplification of the
 CC invention.

XX
 SQ Sequence 20 BP; 5 A; 5 C; 4 G; 6 T; 0 other;

Query Match 12.2%; Score 15; DB 22; Length 20;

Best Local Similarity 100.0%; Pred. No. 97;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 72 ggttaagtccacatg 86
 |||||||||||||
 Db 20 GGTAAGTTCACATG 6

RESULT 28

AAV72858
 ID AAV72858 standard; DNA; 96 BP.

AC AAV72858;

DT 22-FEB-1999 (first entry)

DE Tryptophan 2,3-dioxygenase oligonucleotide variant.

XX
 XX Human: tryptophan 2,3-dioxygenase; TDO2; allelic polygene diagnosis;
 KW reward deficiency syndrome; RDS; obesity; alcohol; tobacco; drug use;
 KW Tourette syndrome; attention deficit hyperactivity disorder; ADHD;
 KW schizoid/avoidant behaviour; aggression; premenstrual syndrome;
 KW violence; hostility; mania; depression; anxiety; sleep problem; autism;
 KW osteoporosis; binge eating; craving; inhibition; PCR primer; ss.

OS Synthetic.

OS Homo sapiens.

PN WO9848785-A2.

PD 05-NOV-1998.

PF 29-APR-1998; 98WO-US08684.

PR 29-APR-1997; 97US-0044394.

XX
 XX (BLUM-) BLUM INC KENNETH.
 PA (CITY) CITY OF HOPE NAT MEDICAL CENT.
 PA (TEXA) UNIV TEXAS SYSTEM.

XX Blum K, Comings DE, Ivy JL;

XX
 DR WPI; 1998-610008/51.
 XX
 PT Composition for treating reward deficiency syndrome behaviour,
 PT attention deficit disorders or controlling weight - contains
 PT inhibitor of opiate peptide destruction, neurotransmitter precursor
 PT and chromium

XX Example 10; Page 293; 663pp; English.

XX A composition (A) has been developed for: (i) treating a reward
 CC deficiency syndrome (RDS) behaviour; or (ii) preventing or treating
 CC unwanted weight gain. (A) comprises: (a) an agent (T), i.e. amino acid
 CC or peptide, or their analogues or derivatives, that inhibits enzymatic
 CC destruction of a neuropeptidyl opiate; (b) a neurotransmitter precursor
 CC (II) to promote neurotransmitter synthesis, i.e. L-tyr, L-phe or L-dopa
 CC (dopamine precursors) L-tyr or 5-hydroxytryptophan (serotonin
 CC precursors); and (C) chromium picolinate or nicotinate to increase the
 CC level of Trp. Typical of many behaviours that can be treated with (A)
 CC include: substance use disorders; obesity; alcohol, tobacco or other
 CC drug use; Tourette syndrome; attention deficit hyperactivity disorder
 CC (ADHD); schizoid/avoidant behaviour; aggression; premenstrual syndrome;
 CC violence; hostility; mania; depression; anxiety; sleep problems; autism.
 CC Methods given in the invention can be applied for diagnosis of RDS or
 CC e.g. elevated levels of low density lipoprotein (LDL) or cholesterol,
 CC longevity, lack of ADHD, osteoporosis. In treatment of obesity, (A)
 CC inhibits binge eating and craving. The present sequence represents an
 CC oligonucleotide used in an example from the present invention.

SQ Sequence 96 BP; 38 A; 7 C; 20 G; 31 T; 0 other;

Query Match 12.2%; Score 15; DB 19; Length 96;

Best Local Similarity 100.0%; Pred. No. 94;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 ggtatcgttaaatgtg 17
 |||||||||||||
 Db 15 ggtatcgttaaatgtg 29

RESULT 29

AAC65347
 ID AAC65347 standard; DNA; 96 BP.

AC AAC65347;

DT 12-FEB-2001 (first entry)

DE Human TDO2 gene G to T variant oligomer #1.

XX
 XX Human: anti-ADD; anti-ADHD; opiate destruction inhibitor;
 KW neurotransmitter synthesis promoter; tryptophan-concentration enhancing;
 KW allelic polygene diagnosis; reward deficiency syndrome; RDS; obesity;
 KW smoking; Tourette's syndrome; schizoid avoidant behaviour; aggression;
 KW post-traumatic stress syndrome; pre-menstrual syndrome;
 KW cannabinoid receptor gene; dopamine DMD4 receptor;
 KW tryptophan 2,3-dioxygenase; TDO2; PCR primer; ss.

OS Homo sapiens.

PN US6132724-A.

PD 17-OCT-2000.

PF 29-APR-1998; 98US-0069886.

XX
 XX (CITY) CITY OF HOPE NAT MEDICAL CENT.
 PA (TEXA) UNIV TEXAS SYSTEM.

XX

PI Blum K;
XX WPI: 2000-678662/66.
DR
XX
PT Compositions used to treat attention deficit (hyperactivity) disorder
PT have of neuroleptidyl opiate enzymatic destruction inhibitor.
PT neurotransmitter, mineral compound and Rhodiola extract or huperzine
XX
XX Example 10; Column 169; 207pp; English.
XX
CC The present sequence is used in a method for the allelic polygene
CC diagnosis of reward deficiency syndrome. Compositions are disclosed which
CC comprise at least one substance that inhibits the enzymatic destruction
CC of a neuroleptidyl opiate, at least one neurotransmitter, a tryptophan
CC concentration enhancing amount of a mineral compound; and at least one
CC substance chosen from Rhodiola extract and huperzine. The compositions
CC are used to treat ADD and ADHD. They are used to treat reward deficiency
CC syndrome (RDS) behaviours including obesity, smoking, Tourette's
CC syndrome, schizoid/avoidant behaviour, aggression, post-traumatic stress
CC syndrome, pre-menstrual syndrome or tobacco use. The presence of
CC encephalin releasers dramatically improves the patient's response to
CC treatment.
XX
SQ Sequence 96 BP; 38 A; 7 C; 20 G; 31 T; 0 other;

Query Match 12.2%; Score 15; DB 21; Length 96;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 ggtattgtaaatgtg 17
|||||
DB 15 ggtattgtaaatgtg 29

RESULT 30
AAV72819
ID AAV72819 standard; DNA; 101 BP.
XX
AC AAV72819;
XX
XX 22-FEB-1999 (first entry)
XX
DE Tryptophan 2,3-dioxygenase oligonucleotide SEQ ID NO:27.
XX
XX Human; tryptophan 2,3-dioxygenase; TDO2; allelic polygene diagnosis;
KW reward deficiency syndrome; RDS; obesity; alcohol; tobacco; drug use;
KW Tourette syndrome; attention deficit hyperactivity disorder; ADHD;
KW schizoid/avoidant behaviour; aggression; premenstrual syndrome;
KW violence; hostility; mania; depression; anxiety; sleep problem; autism;
KW osteoporosis; binge eating; craving; inhibition; PCR primer; ss.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN W09848785-A2.
XX
PD 05-NOV-1998.
XX
PE 29-APR-1998; 98WO-US08684.
XX
PR 29-APR-1997; 97US-0044394.
XX
XX (BLUM-) BLUM INC KENNETH.
PA (CITY) CITY OF HOPE NAT MEDICAL CENT.
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Blum K, Comings DE, Ivy JL;
XX
XX WPI: 1998-610008/51.
XX
PT Composition for treating reward deficiency syndrome behaviour,
PT attention deficit disorders or controlling weight - contains

PT inhibitor of opiate peptide destruction, neurotransmitter precursor
PT and Chromium
XX
XX Example 19; Page 398; 663pp; English.
XX
CC A composition (A) has been developed for: (i) treating a reward
CC deficiency syndrome (RDS) behaviour; or (ii) preventing or treating
CC unwanted weight gain. (A) comprises: (a) an agent (I), i.e. amino acid
CC or peptide, or their analogues or derivatives, that inhibits enzymatic
CC destruction of a neuroleptidyl opiate; (b) a neurotransmitter precursor
CC (II) to promote neurotransmitter synthesis, i.e. L-Tyr, L-Phe or L-dopa
CC (dopamine precursors) L-Tyr or 5-hydroxytryptophan (serotonin
CC precursors) or L-Glu (or its salt) or L-Gln (gamma-aminobutyric acid
CC precursors); and (c) chromium picolinate or nicotinate to increase the
CC level of Trp. Typical of many behaviours that can be treated with (A)
CC include: substance use disorders; obesity; alcohol, tobacco or other
CC drug use; Tourette syndrome; attention deficit hyperactivity disorder
CC (ADHD); schizoid/avoidant behaviour; aggression; premenstrual syndrome;
CC violence; hostility; mania; depression; anxiety; sleep problems; autism.
CC Methods given in the invention can be applied for diagnosis of RDS or
CC e.g. elevated levels of low density lipoprotein (LDL) or cholesterol,
CC longevity, lack of ADHD, osteoporosis. In treatment of obesity, (A)
CC inhibits binge eating and craving. The present sequence represents an
CC oligonucleotide used in an example from the present invention.
XX
SQ Sequence 101 BP; 40 A; 7 C; 22 G; 32 T; 0 other;

Query Match 12.2%; Score 15; DB 19; Length 101;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 ggtattgtaaatgtg 17
|||||
DB 15 ggtattgtaaatgtg 29

RESULT 31
AAC65362
ID AAC65362 standard; DNA; 101 BP.
XX
AC AAC65362;
XX
XX 12-FEB-2001 (first entry)
XX
DE Human TDO2 gene G to T variant oligomer #4.
XX
XX Human; anti-ADD; anti-ADHD; opiate destruction inhibitor;
KW neurotransmitter synthesis promoter; tryptophan-concentration enhancing;
KW allelic polygene diagnosis; reward deficiency syndrome; RDS; obesity;
KW smoking; Tourette's syndrome; schizoid/avoidant behaviour; aggression;
KW post-traumatic stress syndrome; pre-menstrual syndrome;
KW cannabinoid receptor gene; dopamine DRD4 receptor;
KW tryptophan 2,3-dioxygenase; TDO2; PCR primer; ss.
XX
OS Homo sapiens.
XX
PN US6132724-A.
XX
PD 17-OCT-2000.
XX
PE 29-APR-1998; 98US-0069886.
XX
PR 29-APR-1998; 98US-0069886.
XX
XX (CITY) CITY OF HOPE NAT MEDICAL CENT.
PA (TEXA) UNIV TEXAS SYSTEM.
XX
PI Blum K;
XX
XX WPI: 2000-678662/66.
XX
PT Compositions used to treat attention deficit (hyperactivity) disorder

PT have of neuropeptidyl opiate enzymatic destruction inhibitor,
CC neurotransmitter, mineral compound and Rhodiola extract or huperzine
PS Example 19; Column 233; 207pp; English.
XX
CC The present sequence is used in a method for the allelic polygene
CC diagnosis of reward deficiency syndrome. Compositions are disclosed which
CC comprise at least one substance that inhibits the enzymatic destruction
CC of a neuropeptidyl opiate, at least one neurotransmitter, a tryptophan
CC concentration-enhancing amount of a mineral compound; and at least one
CC substance chosen from Rhodiola extract and huperzine. The compositions
CC are used to treat ADD and ADHD. They are used to treat reward deficiency
CC syndrome (RDS) behaviours including obesity, smoking, Tourette's
CC syndrome, schizoid avoidant behaviour, aggression, post-traumatic stress
CC syndrome, pre-menstrual syndrome or tobacco use. The presence of
CC enkephalin releasers dramatically improves the patient's response to
CC treatment.
XX
SQ Sequence 101 BP; 40 A; 7 C; 22 G; 32 T; 0 other;

Query Match 12.2%; Score 15; DB 21; Length 101;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 ggtattgtaaatgt 17
| | | | | | | | | | | | | | | | | | | | | |
Db 15 ggtattgtaaatgt 29

RESULT 32
AAZ36299
ID AAZ36299 standard; cDNA; 201 BP.
XX
AC AAZ36299;
XX
DT 22-FEB-2000 (first entry)
XX
DE Mechanical stress induced [glutamy]-cystein synthetase cDNA.
XX
KW Expressed sequence tag; EST; mechanical stress; gene therapy;
KM osteoporosis; bone density; bone development; ss.
XX
OS Rattus sp.
XX
PN WO960164-A1.
XX
PD 25-NOV-1999.
XX
PF 14-MAY-1999; 99WO-US11066.
XX
PR 15-MAY-1998; 98US-0085673.
XX
PA (OVAR-) QDARK BIOTECH INC.
XX
PI Elnat P, Mor O, Skalter R, Feinstein E, Faerman A;
XX
DR WPI; 2000-053304/04.
XX
PT Identification of stress induced genes for determining risk and
XX preventing, treating or controlling osteoporosis
XX
PS Claim 25; Fig 2A-5; 308pp; English.
XX
CC AAZ36299-236320 represent genes or expressed sequence tags (ESTs)
CC identified by the method of the invention. The sequences were isolated
CC from rat osteoblasts. The specification describes a method for the
CC identification of genes responsive to a specific mechanical stress. The
CC method comprises applying the mechanical stress to an organism (tissue
CC or cells comprising bone cells), isolating the specific cellular
CC fractions and extracting mRNA from them, and differentially analysing the
CC mRNA in comparison with control samples. The method is used to identify
CC genes whose expression is responsive to a specific stress. The identified

CC genes are employed in determining risk associated with a physiological
CC or disease state. The risk determination methods are used for testing a
CC medicament for gene therapy. These medicaments, or genes identified by
CC the method of the invention, are used for treating, preventing or
CC controlling a physiological or disease state (especially osteoporosis
CC or bone density or other factors causing or contributing to osteoporosis
CC or its symptoms or other conditions involved in mechanical stress or its
CC lack. The methods can also be used for advancing research or studies in
CC bone development.
XX
SQ Sequence 201 BP; 56 A; 42 C; 50 G; 53 T; 0 other;

Query Match 12.2%; Score 15; DB 21; Length 201;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 aggtattgtaaatgt 16
| | | | | | | | | | | | | | | | | | | | | |
Db 121 aggtattgtaaatgt 135

RESULT 33
AAC13255/C
ID AAC13255 standard; cDNA; 203 BP.
XX
AC AAC13255;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 17330.
XX
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KM gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI; 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
XX obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
XX diagnostic, forensic, gene therapy and chromosome mapping procedures
XX
PS Claim 1; SEQ ID 17330; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 203 BP; 57 A; 44 C; 35 G; 59 T; 8 other;

Query Match 12.2%; Score 15; DB 21; Length 203;
 Best Local Similarity 100.0%; Pred. No. 92;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 77 gttcacatgtaag 91
 |||||||
 Db 15 GTTTCACATGTAAG 1

RESULT 34

AAA01492
 ID AAA01492 standard; cDNA; 280 BP.

AC AAA01492;

DT 19-MAY-2000 (first entry)

DE Human colon cancer cell line polynucleotide sequence SEQ ID NO:1483.

XX Human; colon cancer; tumour; diagnosis; gene expression product;

KM probe; detection; cancerous state; metastasis; identification;

KW breast cancer; oestrogen receptor-positive breast cancer; therapy;

XX oestrogen receptor-negative breast cancer; lung cancer; ss.

OS Homo sapiens.

PN MO9958675-A2.

PD 18-NOV-1999.

PE 13-MAY-1999; 99WO-US10602.

PR 14-MAY-1998; 98US-0085426.

PR 15-MAY-1998; 98US-0085537.

PR 21-OCT-1998; 98US-0085696.

PR 27-OCT-1998; 98US-0105234.

PR 27-OCT-1998; 98US-0105877.

PR 27-OCT-1998; 98US-0105877.

PR 27-OCT-1998; 98US-0105877.

PR 27-OCT-1998; 98US-0105877.

PR 27-OCT-1998; 98US-0105877.

PR 27-OCT-1998; 98US-0105877.

PR 27-OCT-1998; 98US-0105877.

PR 27-OCT-1998; 98US-0105877.

PR 27-OCT-1998; 98US-0105877.

PR 27-OCT-1998; 98US-0105877.

PR 27-OCT-1998; 98US-0105877.

PR 27-OCT-1998; 98US-0105877.

PR 27-OCT-1998; 98US-0105877.

PR 27-OCT-1998; 98US-0105877.

PR 27-OCT-1998; 98US-0105877.

PR 27-OCT-1998; 98US-0105877.

PR 27-OCT-1998; 98US-0105877.

Sequence 280 BP; 63 A; 45 C; 51 G; 107 T; 14 other;

Query Match 12.2%; Score 15; DB 21; Length 280;
 Best Local Similarity 100.0%; Pred. No. 92;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 104 tcaattattccaca 118
 |||||||
 Db 83 tcaattattccaca 97

RESULT 35

AAZ14187/C
 ID AAZ14187 standard; cDNA; 300 BP.

AC AAZ14187;

DT 12-OCT-1999 (first entry)

DE Human gene expression product cDNA sequence SEQ ID NO:1656.

XX Human; gene; gene expression product; diagnosis; therapy; probe;

KM detection; mapping; tissue typing; profiling; forensic; cancer;

KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.

OS Homo sapiens.

PN MO9938972-A2.

PD 05-AUG-1999.

PE 28-JAN-1999; 99WO-US01619.

PR 03-APR-1998; 98US-0080666.

PR 28-JAN-1998; 98US-0072910.

PR 24-FEB-1998; 98US-0075954.

PR 31-MAR-1998; 98US-0080114.

PR 03-APR-1998; 98US-0080515.

PR 03-APR-1998; 98US-0080515.

PR 03-APR-1998; 98US-0080515.

PR 03-APR-1998; 98US-0080515.

PR 03-APR-1998; 98US-0080515.

PR 03-APR-1998; 98US-0080515.

PR 03-APR-1998; 98US-0080515.

PR 03-APR-1998; 98US-0080515.

PR 03-APR-1998; 98US-0080515.

PR 03-APR-1998; 98US-0080515.

PR 03-APR-1998; 98US-0080515.

PR 03-APR-1998; 98US-0080515.

PR 03-APR-1998; 98US-0080515.

PR 03-APR-1998; 98US-0080515.

PR 03-APR-1998; 98US-0080515.

PR 03-APR-1998; 98US-0080515.

PR 03-APR-1998; 98US-0080515.

Novel human genes and their expression products which are
 differentially expressed in different cell types

Claim 1; Page 1035; 2479pp; English.

WPI; 1999-494092/41.

Novel human genes and their expression products which are

differentially expressed in different cell types

Claim 1; Page 1035; 2479pp; English.

The present invention describes a library of human polynucleotides

comprising the sequences given in AAZ12532 to AAZ1779. Also described is

a method of detecting differentially expressed genes correlated with the

cancerous state of a mammalian cell, comprising detecting at least one

differentially expressed gene product in a test sample from a cell

suspected of being cancerous, where the gene product is encoded by one

of the 5248 polynucleotide sequences given in AAZ12532 to AAZ1779. The

polynucleotides can be used as a source of primers and probes, which can

be used for a variety of purpose, e.g. detection of expression levels,

mapping, tissue typing or profiling, forensics, genetic analysis and

detection of polymorphisms. Polypeptides encoded by the polynucleotides

can be used for raising antibodies for experimental, diagnostic and

therapeutic purposes. The polynucleotides may also be used to construct

arrays for diagnostics (which may be used to determine function of an

encoded protein); and to detect differences in expression levels between

two cells (e.g. to identify abnormal or diseased tissue in a human, to

identify a genetic predisposition or susceptibility to a disease such as

Sequence 303 BP; 84 A; 80 C; 69 G; 70 T; 0 other;

CC AAC7444 to AAC7606 encode the proteins given in ABA43397
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC anisporatic; antiparkinsonian; nootropic; neuroprotective;
CC osteoplastic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;
CC antithyroid; and antianemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy

CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, anti-inflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 309 BP; 89 A; 81 C; 59 G; 80 T; 0 other;

Query Match 12.2%; Score 15; DB 21; Length 309;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 8 tgttaattgagatg 22
Db 224 TGTAAATGTGAGATG 210

RESULT 38
AAK40689/c
ID AAK40689 standard; cDNA; 319 BP.
XX
AC AAK40689;
XX
DT 18-JUN-1999 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 289.
XX
KW Human: secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide; prostate;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS Homo sapiens.
XX
PN W09906550-A2.
XX
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01232.
XX
PR 01-AUG-1997; 97US-0905144.
XX
PA (GEST) GENSET.
XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI: 1999-153780/13.
DR P-PSDB: AAY11967.
XX
PT New isolated prostate-derived nucleic acids - used to develop
PT products which may have cytokine, immune regulatory, haematopoiesis
PT regulating, anti-inflammatory or tumour inhibition activity
XX
XX Claim 1; Page 453; 675pp; English.
XX
CC AAK40689 to AAK40715 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins expressed in prostate, and encode the proteins
CC given in AAY11716 to AAY11933 respectively. The proteins given represent
CC the signal peptide and an N-terminal fragment of a secreted protein. The
CC nucleic acid sequences can be used for producing secreted human gene
CC products. They can also be used to develop products for diagnosis and
CC therapy. The proteins obtained may have cytokine activity, cell
CC proliferation and differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory

CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptides can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ Sequence 319 BP; 61 A; 83 C; 110 G; 58 T; 7 other;

Query Match 12.2%; Score 15; DB 20; Length 319;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 56 gagagctgactgac 70
Db 112 GAGGCGTCGACCTGAC 98

RESULT 39
AAC22215/c
ID AAC22215 standard; cDNA; 357 BP.
XX
AC AAC22215;
XX
DT 06-OCT-2000 (first entry)
XX
DE Human secreted protein 5' EST, SEQ ID NO: 26290.
XX
KW Human: 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX
OS Homo sapiens.
XX
PN EP1033401-A2.
XX
PD 06-SEP-2000.
XX
PF 21-FEB-2000; 2000EP-0200610.
XX
PR 26-FEB-1999; 99US-0122487.
XX
PA (GEST) GENSET.
XX
PI Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR WPI: 2000-500381/45.
XX
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX Claim 1; SEQ ID 26290; 71pp + CD-ROM; English.
XX
CC The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC identified within the present sequence. The 5' ESTs were prepared from
CC total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC sequences usually correspond mainly to the 3' untranslated region (UTR)
CC of the mRNA because they are often obtained from oligo-dT primed cDNA
CC libraries. Such ESTs are not well suited for isolating cDNA sequences
CC derived from the 5' ends of mRNAs and even in those cases where longer
CC cDNA sequences have been obtained, the full 5' UTR is rarely included.
CC 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be
CC used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used
CC in diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC They are used to obtain upstream regulatory sequences and to design
CC expression and secretion vectors.
XX
SQ Sequence 357 BP; 118 A; 56 C; 68 G; 111 T; 4 other;

Query Match 12.2%; Score 15; DB 21; Length 357;

Best Local Similarity 100.0%; Pred. No. 91;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 98 ttatctcaatttat 112
|||||
Db 202 ttatcttCAATTtAT 188

RESULT 40

AAf66728
ID AAF66728 standard; cDNA: 377 BP.

AC AAF66728;

DT 09-APR-2001 (first entry)

DE Novel human polynucleotide, SEQ ID NO: 2484.

KM Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW breast cancer; lung cancer; cancer detection; ss.

OS Homo sapiens.

PN WO200102568-A2.

PD 11-JAN-2001.

PF 30-JUN-2000; 2000WO-US18374.

PR 02-JUL-1999; 99US-0142310.

PR 02-JUL-1999; 99US-0142311.

PA (CHIR) CHIRON CORP.
(HYSE-) HYSEQ INC.

PI Williams LR, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;

PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;

PI Kita D, Garcia V, Jones LW, Strache-Crain B;

DR WPI; 2001-091805/10.

PT Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -

PS Claim 9; Page 910; 1046pp; English.

CC The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia.

XX Sequence 377 BP; 116 A; 57 C; 76 G; 128 T; 0 other;

Query Match 12.2%; Score 15; DB 22; Length 377;

Best Local Similarity 100.0%; Pred. No. 91;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 ggtatgttaaatgtg 17
|||||

Db 20 ggtatgttaaatgtg 34

RESULT 41

AAf66301/C
ID AAF66301 standard; cDNA: 391 BP.

AC AAF66301;

DT 09-APR-2001 (first entry)

DE Novel human polynucleotide, SEQ ID NO: 2057.

KM Human; cytostatic; gene therapy; colon cancer; prostate cancer;
KW breast cancer; lung cancer; cancer detection; ss.

OS Homo sapiens.

PN WO200102568-A2.

PD 11-JAN-2001.

PF 30-JUN-2000; 2000WO-US18374.

PR 02-JUL-1999; 99US-0142310.

PR 02-JUL-1999; 99US-0142311.

PA (CHIR) CHIRON CORP.
(HYSE-) HYSEQ INC.

PI Williams LR, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;

PI Crkenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;

PI Kita D, Garcia V, Jones LW, Strache-Crain B;

DR WPI; 2001-091805/10.

PT Library of polynucleotides for diagnosing a cancerous state of a
PT mammalian cell and detecting cancer, particularly of the colon or
PT prostate, comprises 3351 human polynucleotide sequences -

PS Claim 9; Page 842; 1046pp; English.

CC The present sequence is one of 3351 sequences in a library of human
CC polynucleotides. The library is used to detect differentially expressed
CC genes correlated with a cancerous state of a mammalian cell and can
CC detect colon, prostate, breast and lung cancer. The library can be used
CC to produce probes for detection of mRNA and to produce additional copies
CC of the polynucleotides. The probes can be used for chromosome mapping of
CC the polynucleotide and for detection of transcription levels. Ribozymes
CC or antisense oligonucleotides can be generated. The polynucleotides and
CC their gene products are used as genetic or biochemical markers (e.g. in
CC blood or tissues) that will detect the earliest changes along the
CC carcinogenesis pathway and/or monitor the efficacy of therapies and
CC preventive interventions. The polynucleotides, polypeptides and
CC antibodies against them can be used in pharmaceutical compositions to
CC treat the cancers and proliferative disorders such as neoplasia,
CC dysplasia and hyperplasia.

XX Sequence 391 BP; 85 A; 105 C; 130 G; 71 T; 0 other;

Query Match 12.2%; Score 15; DB 22; Length 391;

Best Local Similarity 100.0%; Pred. No. 91;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 gaagctgcactgac 70
|||||

Db 105 GAAGCTGCACtGAC 91

RESULT 42
AAf66405/C

```

ID  AAF66405 standard; cDNA; 412 BP.
XX
XX  AAF66405;
AC
XX  09-APR-2001 (first entry)
DT
XX
XX  Novel human polynucleotide, SEQ ID NO: 2161.
DE
XX
XX  Human; cytosolic; gene therapy; colon cancer; prostate cancer;
KW  breast cancer; lung cancer; cancer detection; ss.
XX
OS  Homo sapiens.
XX
XX  W0200102568-A2.
XX
XX  11-JAN-2001.
XX
XX  30-JUN-2000; 2000WO-US18374.
XX
XX  02-JUL-1999; 99US-0142310.
XX  02-JUL-1999; 99US-0142311.
XX
XX  (CHIR ) CHIRON CORP.
PA  (HYSE-) HYSEQ INC.
XX
XX  Williams LT, Escobedo J, Innis MA, Garcia PD, Klinger J, Kassam A;
PI  Reinhard C, Randazzo F, Kennedy GC, Pot D, Lamson G, Drmanac R;
PI  Cifrenjakov R, Drmanac S, Dickson M, Labat I, Leshkowitz D;
PI  Kita D, Garcia V, Jones LW, Strache-Crain B;
XX
XX  WPI: 2001-091805/10.
DR
XX
XX  Library of polynucleotides for diagnosing a cancerous state of a
PT  mammalian cell and detecting cancer, particularly of the colon or
PT  prostate, comprises 3351 human polynucleotide sequences -
XX
XX  Claim 9; Page 859; 1046pp; English.
XX
XX  The present sequence is one of 3351 sequences in a library of human
CC  polynucleotides. The library is used to detect differentially expressed
CC  genes correlated with a cancerous state of a mammalian cell and can
CC  detect colon, prostate, breast and lung cancer. The library can be used
CC  to produce probes for detection of mRNA and to produce additional copies
CC  of the polynucleotides. The probes can be used for chromosome mapping of
CC  the polynucleotide and for detection of transcription levels. Ribozymes
CC  or antisense oligonucleotides can be generated. The polynucleotides and
CC  their gene products are used as genetic or biochemical markers (e.g. in
CC  blood or tissues) that will detect the earliest changes along the
CC  carcinogenesis pathway and/or monitor the efficacy of therapies and
CC  preventive interventions. The polynucleotides, polypeptides and
CC  antibodies against them can be used in pharmaceutical compositions to
CC  treat the cancers and proliferative disorders such as neoplasia,
CC  dysplasia and hyperplasia.
XX
XX  Sequence 412 BP; 113 A; 99 C; 85 G; 115 T; 0 other;
SQ

```

Query Match 12.2%; Score 15; DB 22; Length 412;
 Best Local Similarity 100.0%; Pred. No. 91;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY  60 gctgcactgactggt 74
    |||||||
DB  333 GCTGCACGTGACTGCT 319

```

RESULT 43
 AAC01762/c
 ID AAC01762 standard; cDNA; 434 BP.
 XX
 AC AAC01762;
 XX
 DT 06-OCT-2000 (first entry)

```

XX
XX  Human secreted protein 5' EST, SEQ ID NO: 1760.
DE
XX
XX  Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW  gene therapy; chromosome mapping; ss.
XX
XX  Homo sapiens.
OS
XX
XX  EP1033401-A2.
XX
XX  06-SEP-2000.
XX
XX  21-FEB-2000; 2000EP-0200610.
XX
XX  26-FEB-1999; 99US-0122487.
XX
XX  (GSEST ) GENSET.
XX
XX  Dumas Milne Edwards J, Duclert A, Giordano J;
PI
XX
XX  WPI: 2000-500381/45.
XX  P-PSDB; AAG01756.
DR
XX
XX  New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT  obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT  diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
XX  Claim 1; SEQ ID 1760; 71pp + CD-ROM; English.
XX
XX  The present sequence is one of a large number of 5' ESTs derived from
CC  mRNAs encoding secreted proteins. An ORF has been identified within the
CC  sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC  derived from 30 different tissues. EST sequences usually correspond
CC  mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC  often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC  well suited for isolating cDNA sequences derived from the 5' ends of
CC  mRNAs and even in those cases where longer cDNA sequences have been
CC  obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC  mRNAs with intact 5' ends and can therefore be used to obtain full length
CC  cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC  gene therapy and chromosome mapping procedures. They are used to obtain
CC  upstream regulatory sequences and to design expression and secretion
CC  vectors.
XX
XX  Sequence 434 BP; 123 A; 116 C; 124 G; 70 T; 1 other;
SQ

```

Query Match 12.2%; Score 15; DB 21; Length 434;
 Best Local Similarity 100.0%; Pred. No. 91;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY  87 taagtgatcattat 101
    |||||||
DB  375 TAAAGTGATCATTAT 361

```

RESULT 44
 AAF10271
 ID AAF10271 standard; cDNA; 556 BP.
 XX
 AC AAF10271;
 XX
 DT 13-MAR-2001 (first entry)
 DE
 DE Fusarium venenatum EST SEQ ID NO:2794.
 XX
 XX Multiple gene expression; filamentous fungal cell; EST;
 KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
 KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
 KW culture condition; environmental stress; spore morphogenesis;
 KW metabolic pathway engineering; catabolic pathway engineering; ss.
 XX
 OS Fusarium venenatum.

XX WO200056762-A2.
PN
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
XX (NOVO) NOVO NORDISK AS.
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB,
XX WPI; 2000-594572/56.
DR
XX
PT Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
XX
PS Claim 86; Page 1396; 3161pp; English.
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AA07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 556 BP; 160 A; 128 C; 138 G; 128 T; 2 other;
XX
Query Match 12.2%; Score 15; DB 21; Length 556;
Best Local Similarity 100.0%; Pred No. 90;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 43 aatgagcgaatgcga 57
XXXXXXXXXXXXXXXXXXXX
Db 191 aatgagcgaatgcga 205
XX
RESULT 45
AAC87998
ID AAC87998 standard; DNA; 567 BP.
XX
AC AAC87998;
XX
DT 07-MAR-2001 (first entry)
XX
XX Human CLASP-2 genomic DNA from BAC ref 13.1.
DE
XX
XX CLASP-1; CLASP-2; transmembrane protein; immune response; inflammatory;
KW cadherin-like asymmetry protein; autoimmune disease; immunosuppressive;
KW immunomodulatory; antiinflammatory; antitumor; cytostatic;
KW hypotensive; antirheumatic; antinaemic; haemostatic; neuroprotective;

KW hypersensitivity; transplantation rejection response; immunodeficiency;
KW proliferation; differentiation; inflammatory response; arthritis;
KW inflammatory bowel disease; haematopoietic cell; blood protein disorder;
KW anaemia; thrombocytopenia; multiple sclerosis; rheumatoid arthritis;
KW endometriosis; pregnancy induced hypertension; ds.
XX
OS Homo sapiens.
XX
PN WO200061747-A2.
XX
PD 19-OCT-2000.
XX
PF 13-APR-2000; 2000WO-US10158.
XX
XX
PR 14-APR-1999; 99US-0129171.
PR 14-MAY-1999; 99US-0134114.
PR 14-MAY-1999; 99US-0134117.
PR 14-MAY-1999; 99US-0134118.
PR 21-OCT-1999; 99US-0160860.
PR 29-OCT-1999; 99US-0162498.
PR 13-DEC-1999; 99US-0170453.
PR 14-JAN-2000; 2000US-0176195.
PR 14-FEB-2000; 2000US-0182296.
XX
XX (ARBO-) ARBOR VITA CORP.
XX
PI Lu PS;
XX
DR WPI; 2000-619230/59.
XX
PT Isolated cadherin-like asymmetry protein-2 polynucleotide and
PT polypeptide used to diagnose, treat and prevent autoimmune diseases and
PT inflammatory responses -
XX
XX
PS Example 3; Fig 6A; 286pp; English.
XX
XX The present invention describes cadherin-like asymmetry protein-2
XX (CLASP-2). CLASP-2 can have immunosuppressive, immunomodulatory,
XX antiinflammatory, antitumor, cytostatic, hypotensive, antirheumatic,
XX antinaemic, haemostatic and neuroprotective activities. CLASP-2 can be
XX used to inhibit an immune response in a subject by interfering with the
XX ability of a CLASP-2 protein to bind to another T cell or B cell. An
XX immune response in a subject may also be inhibited by administering an
XX antibody which specifically binds to CLASP-2. CLASP-2 polynucleotides,
XX proteins and antibodies can be used to prevent or treat a CLASP-2
XX mediated disease, such as an autoimmune disease caused or exacerbated
XX by increased activity of Th1 cells. They can also be used to treat
XX hypersensitivities, prevent transplantation rejection responses and
XX augment immune responsiveness in immunodeficiency states, inhibit
XX proliferation and differentiation of cells involved in an inflammatory
XX response e.g. arthritis, inflammatory bowel disease and increase
XX differentiation and proliferation of haematopoietic cells e.g. to treat
XX anaemia, thrombocytopenia and other blood protein disorders. Disorders
XX treated by disrupting CLASP-2 function include multiple sclerosis,
XX rheumatoid arthritis, endometriosis and pregnancy induced hypertension.
XX The present sequence represents a human CLASP-2 genomic DNA sequence
XX from a bacterial artificial chromosome (BAC), given in the present
XX invention.
XX
SQ Sequence 567 BP; 161 A; 89 C; 105 G; 188 T; 24 other;
XX
Query Match 12.2%; Score 15; DB 21; Length 567;
Best Local Similarity 100.0%; Pred No. 90;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 56 gagggctgactgac 70
XXXXXXXXXXXXXXXXXXXX
Db 84 gagggctgactgac 98

Job time: 2912 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2002, 10:00:11 ; Search time 39.53 Seconds
(without alignments)
704.700 Million cell updates/sec

Title: US-09-684-016-48411

Perfect score: 123
Sequence: 1 caggtatgttaaatgtgaga.....tcaatttaccacacaca 123

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents, NA:*

1: /cgn2_6/ptodata/2/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/lna/PCrUS.COMB.seq:*
6: /cgn2_6/ptodata/2/lna/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	16	13.0	1278	4 US-09-050-159-123	Sequence 123, App
C 2	15	12.2	96	3 US-09-069-886-12	Sequence 12, Appl
C 3	15	12.2	101	3 US-09-069-886-27	Sequence 27, Appl
C 4	15	12.2	1099	2 US-08-871-033-2	Sequence 2, Appl
C 5	15	12.2	1099	2 US-09-212-167-2	Sequence 2, Appl
C 6	15	12.2	3171	5 PCR-US95-09261-1	Sequence 1, Appl
C 7	14	11.4	528	3 US-09-157-177-111	Sequence 111, App
C 8	14	11.4	564	4 US-09-328-111-223	Sequence 223, App
C 9	14	11.4	574	4 US-09-328-111-297	Sequence 297, App
C 10	14	11.4	750	4 US-08-845-256-2	Sequence 2, Appl
C 11	14	11.4	877	2 US-09-208-210-2	Sequence 2, Appl
C 12	14	11.4	1021	3 US-08-989-289-2	Sequence 2, Appl
C 13	14	11.4	1021	3 US-08-714-918-70	Sequence 70, Appl
C 14	14	11.4	1021	4 US-09-265-315-70	Sequence 70, Appl
C 15	14	11.4	1021	4 US-09-265-315-70	Sequence 70, Appl
C 16	14	11.4	1021	4 US-09-265-315-70	Sequence 70, Appl
C 17	14	11.4	1620	2 US-09-266-417-70	Sequence 70, Appl
C 18	14	11.4	1620	2 US-08-874-186-37	Sequence 37, Appl
C 19	14	11.4	2433	1 US-08-222-289-1	Sequence 1, Appl
C 20	14	11.4	2466	6 US-08-248-021A-1	Sequence 1, Appl
C 21	14	11.4	2466	6 5194375-5	Patent No. 5194375
C 22	14	11.4	4259	3 US-08-816-155B-2	Sequence 2, Appl
C 23	14	11.4	4739	3 US-09-079-587-2	Sequence 2, Appl
C 24	14	11.4	5455	1 US-08-685-871-1	Sequence 1, Appl
C 25	14	11.4	8561	3 US-08-342-930-1	Sequence 3, Appl
C 26	14	11.4	10281	2 US-09-112-450-3	Sequence 1, Appl
C 27	14	11.4	10281	3 US-08-816-155B-1	Sequence 1, Appl
C 27	14	11.4	10281	3 US-09-079-587-1	Sequence 1, Appl

C 28	14	11.4	43795	3 US-08-742-185-101	Sequence 101, App
C 29	13	10.6	30	4 US-09-503-391-16	Sequence 16, Appl
C 30	13	10.6	33	4 US-08-427-569-14	Sequence 14, Appl
C 31	13	10.6	36	1 US-08-311-760A-315	Sequence 315, App
C 32	13	10.6	36	1 US-08-311-760A-315	Sequence 315, App
C 33	13	10.6	36	2 US-08-774-310-315	Sequence 315, App
C 34	13	10.6	36	2 US-08-774-310-315	Sequence 316, App
C 35	13	10.6	39	1 US-08-207-226A-3	Sequence 3, Appl
C 36	13	10.6	40	2 US-08-425-684-133	Sequence 133, App
C 37	13	10.6	40	2 US-08-675-502-133	Sequence 133, App
C 38	13	10.6	42	3 US-09-129-740-10	Sequence 10, Appl
C 39	13	10.6	50	2 US-08-821-559A-17	Sequence 17, Appl
C 40	13	10.6	70	5 PCR-US95-14418-50	Sequence 50, Appl
C 41	13	10.6	70	5 PCR-US95-15327-50	Sequence 50, Appl
C 42	13	10.6	71	4 US-08-569-147-69	Sequence 69, Appl
C 43	13	10.6	95	3 US-09-235-246-21	Sequence 21, Appl
C 44	13	10.6	96	1 US-08-253-877C-12	Sequence 12, Appl
C 45	13	10.6	96	2 US-08-452-164A-12	Sequence 12, Appl

ALIGNMENTS

```
RESULT 1
US-09-050-159-123/c
; Sequence 123, Application US/09050159A
; Patent No. 6197505
; GENERAL INFORMATION:
; APPLICANT: NO. 6197505berg, Leif T
; APPLICANT: Andersson, Maria K
; APPLICANT: Linstrom, Per H
; TITLE OF INVENTION: METHODS FOR ASSESSING CARDIOVASCULAR STATUS AND
; FILE REFERENCE: 1248/10042
; CURRENT APPLICATION NUMBER: US/09/050,159A
; EARLIER FILING DATE: 1998-03-27
; EARLIER APPLICATION NUMBER: 60/042,930
; NUMBER OF SEQ. ID NOS: 133
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 123
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Angiotensinogen, 5' region and exon 1
US-09-050-159-123

Query Match 13.0%; Score 16; DB 4; Length 1278;
Best local Similarity 100.0%; Pred. No. 5.3;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 71 tggtaagttcacatg 86
Db 522 TGCTAGTTCACATG 507

RESULT 2
US-09-069-886-12
; Sequence 12, Application US/09069886
; Patent No. 6132724
; GENERAL INFORMATION:
; APPLICANT: Blum, Kenneth
; APPLICANT: Comings, David E.
; APPLICANT: Ivy, John L.
; TITLE OF INVENTION: ALLELIC POLYGENE DIAGNOSIS OF RHEUMATOID
; TITLE OF INVENTION: DEFICIENCY SYNDROME AND TREATMENT
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
```

STATE: TX
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/069,886
FILING DATE: Concurrently Herewith
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Hodgins, Daniel S.
REGISTRATION NUMBER: 31,026
REFERENCE/DOCKET NUMBER: BLUM:002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512)418-3000
TELEFAX: (512)474-7577
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 96 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-069-886-12

Query Match 12.2%; Score 15; DB 3; Length 96;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 ggtattgtaaatgtg 17
|||||
DB 15 GGTATTGTAATGTC 29

RESULT 3
US-09-069-886-27
Sequence 27, Application US/09069886
Patent No. 6132724
GENERAL INFORMATION:
APPLICANT: Blum, Kenneth
APPLICANT: Comings, David E.
TITLE OF INVENTION: ALLELIC POLYGENE DIAGNOSIS OF REWARD
TITLE OF INVENTION: DEFICIENCY SYNDROME AND TREATMENT
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77210-4433
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/069,886
FILING DATE: Concurrently Herewith
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Hodgins, Daniel S.
REGISTRATION NUMBER: 31,026
REFERENCE/DOCKET NUMBER: BLUM:002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512)418-3000
TELEFAX: (512)474-7577
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:

LENGTH: 101 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-069-886-27

Query Match 12.2%; Score 15; DB 3; Length 101;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 ggtattgtaaatgtg 17
|||||
DB 15 GGTATTGTAATGTC 29

RESULT 4
US-08-871-033-2/c
Sequence 2, Application US/08871033
Patent No. 5869259
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
TITLE OF INVENTION: NEW CARBOXYPEPTIDASE INHIBITOR
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/871,033
FILING DATE: Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0318 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1099 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: PROSTUT09
CLONE: 1649584
US-08-871-033-2

Query Match 12.2%; Score 15; DB 2; Length 1099;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 87 taagtgataactat 101
|||||
DB 307 TAAAGTGATACTTAT 293


```
RESULT 5
US-09-212-167-2/c
; Sequence 2, Application US/09212167
; Patent No. 5998373
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NEW CARBOXYPEPTIDASE INHIBITOR
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/212,167
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/871,033
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0318 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1099 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT09
; CLONE: 1649584
; US-09-212-167-2

Query Match 12.2%; Score 15; DB 2; Length 1099;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 taagtgatacttat 101
|||||
DB 307 TAAGTGATCTTAT 293

RESULT 6
PCT-US95-09261-1/c
; Sequence 1, Application PC/TUS9509261
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT: NAME: BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM
; APPLICANT: STREET: 201 West 7th Street
; APPLICANT: CITY: Austin
; APPLICANT: STATE: Texas
; APPLICANT: COUNTRY: United States of America
; APPLICANT: POSTAL CODE: 78701
; APPLICANT: TELEPHONE NO.: (512)499-4462
; APPLICANT: TELEFAX: (512)499-4523
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE EXPRESSION OF
; A BONE AND PROSTATE DERIVED GROWTH FACTOR
; NUMBER OF SEQUENCES: 5
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/09261
; FILING DATE: CONCURRENTLY HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/283,701
; FILING DATE: 01-AUG-1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: SERTICH, GARY J.
; REGISTRATION NUMBER: 34,430
; REFERENCE/DOCKET NUMBER: UTFC422P--
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 694..2310
; PCT-US95-09261-1

Query Match 12.2%; Score 15; DB 5; Length 3171;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 aattattccacaca 120
|||||
DB 3143 AATTATTCCACACA 3129

RESULT 7
US-09-157-177-111/c
; Sequence 111, Application US/09157177
; Patent No. 6090558
; GENERAL INFORMATION:
; APPLICANT: Butler, John M.
; APPLICANT: LI, Jia
; APPLICANT: Monforte, Joseph A.
; APPLICANT: Becker, Christopher H.
; TITLE OF INVENTION: DNA TYPING BY MASS SPECTROMETRY WITH POLYMORPHIC DNA
; TITLE OF INVENTION: REPEAT MARKERS
; FILE REFERENCE: GETR:017/GETR017P
; CURRENT APPLICATION NUMBER: US/09/157,177
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: 60/059,415
; EARLIER FILING DATE: 1997-09-19
; NUMBER OF SEQ ID NOS: 135
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 111
; LENGTH: 528
; TYPE: DNA
; ORGANISM: Homo sapiens
```

US-09-157-177-111

Query Match 11.4%; Score 14; DB 3; Length 528;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 13 atgtgagatgcctt 26
|||||
Db 46 ATGTGAGATGCCTT 33

RESULT 8

US-09-328-111-223
; Sequence 223, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dertl, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.
; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 223
; LENGTH: 564
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(564)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-223

Query Match 11.4%; Score 14; DB 4; Length 564;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 101 tctcaattatttc 114
|||||
Db 117 tctcaattatttc 130

RESULT 9

US-09-328-111-297
; Sequence 297, Application US/09328111
; Patent No. 6262333
; GENERAL INFORMATION:
; APPLICANT: Endege, Wilson O.
; APPLICANT: Steinmann, Kathleen E.
; APPLICANT: Astle, Jon H.
; APPLICANT: Burgess, Christopher C.
; APPLICANT: Bushnell, Steven E.
; APPLICANT: Carroll III, Eddie
; APPLICANT: Catino, Theodore J.
; APPLICANT: Dertl, Adnan
; APPLICANT: Ford, Donna M.
; APPLICANT: Lewis, Marcia E.

; APPLICANT: Monahan, John E.
; APPLICANT: Schlegel, Robert
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCD-257 (US)
; CURRENT APPLICATION NUMBER: US/09/328,111
; CURRENT FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: US 60/088,801
; EARLIER FILING DATE: 1998-06-10
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 297
; LENGTH: 574
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(574)
; OTHER INFORMATION: n = A,T,C or G
US-09-328-111-297

Query Match 11.4%; Score 14; DB 4; Length 574;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 101 tctcaattatttc 114
|||||
Db 144 tctcaattatttc 157

RESULT 10

US-08-845-256-2
; Sequence 2, Application US/08845256
; Patent No. 5874244
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: NEW HUMAN NMDH DEHYDROGENASE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESS: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/845,256
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0277 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-845-4166
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 750 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BEPI10101

CLONE: 2054787
US-08-845-256-2

Query Match 11.4%; Score 14; DB 2; Length 750;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ttgtaaatgtgaga 20
|||||
Db 641 TTGTAATGTGAGA 654

RESULT 11
US-09-208-210-2
; Sequence 2, Application US/09208210
; Patent No. 6277572
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: NEW HUMAN NADH DEHYDROGENASE
; TITLE OF INVENTION: SUBUNITS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/208,210
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/845,256
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0277 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-845-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 750 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BEP1NOT01
; CLONE: 2054787
; US-09-208-210-2

Query Match 11.4%; Score 14; DB 4; Length 750;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ttgtaaatgtgaga 20
|||||
Db 641 TTGTAATGTGAGA 654

RESULT 12
US-08-989-289-2
; Sequence 2, Application US/08989289
; Patent No. 5968747
; GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.
APPLICANT: Shah, Purvi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: UBIQUITIN-LIKE CONJUGATING PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/989,289
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0439 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 877 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: ADREUT05
CLONE: 2501808
US-08-989-289-2

Query Match 11.4%; Score 14; DB 2; Length 877;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 caacctgatgacc 41
|||||
Db 449 CAACCTGATGACC 462

RESULT 13
US-08-714-918-70/C
; Sequence 70, Application US/08714918
; Patent No. 6037123
; GENERAL INFORMATION:
; APPLICANT: Benton, Bret
; APPLICANT: Lee, Ying
; APPLICANT: Malouin, Francois
; APPLICANT: Martin, Patrick K.
; APPLICANT: Schmidt, Molly B.
; APPLICANT: Sun, Dongxu
; TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS ANTIBACTERIAL
; TITLE OF INVENTION: TARGET GENES
; NUMBER OF SEQUENCES: 111
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; CITY: Los Angeles
; STATE: California

COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/714,918
FILING DATE: September 13, 1996
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 222/005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ. ID NO: 70:
SEQUENCE CHARACTERISTICS:
LENGTH: 1021 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-714-918-70

Query Match 11.4%; Score 14; DB 3; Length 1021;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 ggtattgtaattgt 16
|||||
Db 523 GGTATTGTAATGT 510

RESULT 14
US-09-265-315-70/c
Sequence 70, Application US/09265315
Patent No. 6187541
GENERAL INFORMATION:
APPLICANT: Benton, Bret
APPLICANT: Lee, Ying J.
APPLICANT: Malouin, Francois
APPLICANT: Martin, Patrick K.
APPLICANT: Schmid, Molly B.
APPLICANT: Sun, Dongxu
TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/265,315
FILING DATE: March 9, 1999
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/714,918
FILING DATE: September 13, 1996
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 240/247
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ. ID NO: 70:
SEQUENCE CHARACTERISTICS:
LENGTH: 1021 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-265-315-70

Query Match 11.4%; Score 14; DB 4; Length 1021;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 ggtattgtaattgt 16
|||||
Db 523 GGTATTGTAATGT 510

RESULT 15
US-09-265-315-70/c
Sequence 70, Application US/09265315
Patent No. 6187541
GENERAL INFORMATION:
APPLICANT: Benton, Bret
APPLICANT: Lee, Ying J.
APPLICANT: Malouin, Francois
APPLICANT: Martin, Patrick K.
APPLICANT: Schmid, Molly B.
APPLICANT: Sun, Dongxu
TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/265,315
FILING DATE: March 9, 1999
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/714,918
FILING DATE: September 13, 1996
APPLICATION NUMBER: 60/009,102

FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 240/247
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 70:
SEQUENCE CHARACTERISTICS:
LENGTH: 1021 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-265-315-70

Query Match 11.4%; Score 14; DB 4; Length 1021;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 gqatgttaagt 16
|||||
Db 523 GGTATTGTAATGT 510

RESULT 16
US-09-266-417-70/c
Sequence 70, Application US/09266417
Patent No. 6228588
GENERAL INFORMATION:
APPLICANT: Benton, Bret
APPLICANT: Lee, Ying J.
APPLICANT: Malouin, Francois
APPLICANT: Martin, Patrick K.
APPLICANT: Schmid, Molly B.
APPLICANT: Sun, Dongxu
TITLE OF INVENTION: METHODS OF SCREENING FOR COMPOUNDS
TITLE OF INVENTION: ACTIVE ON STAPHYLOCOCCUS AUREUS
TITLE OF INVENTION: TARGET GENES
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/266,417
FILING DATE: March 9, 1999
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/714,918
FILING DATE: September 13, 1996
APPLICATION NUMBER: 60/009,102
FILING DATE: December 22, 1995
APPLICATION NUMBER: 60/003,798
FILING DATE: September 15, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 240/248

TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 70:
SEQUENCE CHARACTERISTICS:
LENGTH: 1021 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-266-417-70

Query Match 11.4%; Score 14; DB 4; Length 1021;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 gqatgttaagt 16
|||||
Db 523 GGTATTGTAATGT 510

RESULT 17
US-08-874-186-37/c
Sequence 37, Application US/08874186
Patent No. 5989885
GENERAL INFORMATION:
APPLICANT: Teng, David H-F.
APPLICANT: Tavliqian, Sean V.
APPLICANT: Perry III, William L.
APPLICANT: Skolnick, Mark H.
TITLE OF INVENTION: SPECIFIC MUTATIONS OF MAP KINASE KINASE
TITLE OF INVENTION: 4 (MKK4) IN HUMAN TUMOR CELL LINES IDENTIFY IT AS A TUMOR
NUMBER OF SEQUENCES: 96
CORRESPONDENCE ADDRESS:
ADDRESSEE: Venable, Baetjer, Howard & Civiletti, LLP
STREET: 1201 New York Avenue, N.W., Suite 1000
CITY: Washington
STATE: DC
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/874,186
FILING DATE:
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/782,482
FILING DATE: 10-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Saxe, Stephen A.
REGISTRATION NUMBER: 38,609
REFERENCE/DOCKET NUMBER: 24884-121392-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-962-4848
TELEFAX: 202-962-8300
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 1620 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: Intron
LOCATION: 1..1325

FEATURE:
NAME/KEY: exon
LOCATION: 1326..1428
FEATURE:
NAME/KEY: Intron
LOCATION: 1429..1620
US-08-874-186-37

Query Match 11.4%; Score 14; DB 2; Length 1620;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 tttaatgtgagat 21
|||||
DB 269 TGTGAGATGAGAT 256

RESULT 18
US-08-222-289-1/c
Sequence 1, Application US/08222289
Patent No. 5559010
GENERAL INFORMATION:
APPLICANT: Kilonsky, Daniel J
APPLICANT: Destruelle, Monika
TITLE OF INVENTION: NUTRIENT REGULATED GENE EXPRESSION
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHNACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/222,289
FILING DATE: 04-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard Aron
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59767/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2165 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 370..1434
US-08-222-289-1

Query Match 11.4%; Score 14; DB 1; Length 2165;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 14 tttgagatgcctta 27
|||||
DB 475 TGTGAGATGCTTA 462

RESULT 19
US-08-248-021A-1/c
Sequence 1, Application US/08248021A
Patent No. 5648240
GENERAL INFORMATION:
APPLICANT: Hook, Magnus
APPLICANT: Jonsson, Klas
APPLICANT: Patti, Joseph M.
APPLICANT: Gurusiddappa, Sivashankarappa
TITLE OF INVENTION: MHC II ANALOG FROM STAPHYLOCOCCUS AUREUS
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P.O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: United States of America
ZIP: 77210

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/248,021A
FILING DATE: 24-MAY-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: TANK:155
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2433 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-248-021A-1

Query Match 11.4%; Score 14; DB 1; Length 2433;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 tttgagatgcctta 27
|||||
DB 2169 TGTGAGATGCTTA 2156

RESULT 20
5194375-5
Patent No. 5194375
APPLICANT: PARK, LINDA S.; GOODWIN, RAYMOND G.
TITLE OF INVENTION: DNA ENCODING INTERLEUKIN-7 RECEPTORS AND
METHODS OF USE
NUMBER OF SEQUENCES: 9
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/493,588
FILING DATE: 21-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 421,201
FILING DATE: 13-OCT-1989
APPLICATION NUMBER: 366,910
FILING DATE: 15-JUN-1989
SEQ ID NO: 5:
LENGTH: 2466
5194375-5

Query Match 11.4%; Score 14; DB 6; Length 2466;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 ttattccacacaa 121
|||||
Db 593 ttattccacacaa 606

RESULT 21

US-08-816-155B-2/c
; Sequence 2, Application US/08816155B
; Patent No. 5990091
; GENERAL INFORMATION:
; APPLICANT: TARTAGLIA, JAMES
; APPLICANT: COX, WILLIAM I.
; APPLICANT: GETTIG, RUSSELL R.
; APPLICANT: MARTINEZ, HECTOR
; APPLICANT: PAOLETTI, ENZO
; APPLICANT: PINCUS, STEVEN E.
; TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND
; TITLE OF INVENTION: METHODS OF MAKING AND USES THEREOF
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
; STREET: 745 FIFTH AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10151
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/816,155B
; FILING DATE: 12-MAR-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: KOWALSKI, THOMAS J.
; REGISTRATION NUMBER: 32,147
; REFERENCE/DOCKET NUMBER: 454310-2990
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-588-0800
; TELEFAX: 212-588-0500
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4259 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-816-155B-2

Query Match 11.4%; Score 14; DB 2; Length 4259;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 attattccacaca 120
|||||
Db 1759 ATTATTCCACACA 1746

RESULT 22

US-09-079-587-2/c
; Sequence 2, Application US/09079587
; Patent No. 6130066
; GENERAL INFORMATION:
; APPLICANT: TARTAGLIA, JAMES
; APPLICANT: COX, WILLIAM I.
; APPLICANT: GETTIG, RUSSELL R.

APPLICANT: MARTINEZ, HECTOR
APPLICANT: PAOLETTI, ENZO
APPLICANT: PINCUS, STEVEN E.
TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND
TITLE OF INVENTION: METHODS OF MAKING AND USES THEREOF
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESSES:
ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
STREET: 745 FIFTH AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10151

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/079,587
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/816,155
FILING DATE: 12-MAR-1997
ATTORNEY/AGENT INFORMATION:
NAME: KOWALSKI, THOMAS J.
REGISTRATION NUMBER: 32,147
REFERENCE/DOCKET NUMBER: 454310-2990
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-588-0800
TELEFAX: 212-588-0500
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 4259 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-079-587-2

Query Match 11.4%; Score 14; DB 3; Length 4259;
Best Local Similarity 100.0%; Pred. No. 69;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 attattccacaca 120
|||||
Db 1759 ATTATTCCACACA 1746

RESULT 23

US-08-685-871-1
; Sequence 1, Application US/08685871
; Patent No. 6013499
; GENERAL INFORMATION:
; APPLICANT: NARUMIYA, Shuh
; APPLICANT: IWAMATSU, Akihiro
; TITLE OF INVENTION: RHO TARGET PROTEIN KINASE P160
; NUMBER OF SEQUENCES: 68
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/685,871
FILING DATE: 24-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-184102
FILING DATE: 25-JUN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-262553
FILING DATE: 14-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16887/845
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4739 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 448..4509
US-08-685-871-1

Query Match 11.4% Score 14; DB 3; Length 4739;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 93 gactactatctca 106
|||||
Db 1658 GACTACTACTCTCA 1671

RESULT 24
US-08-342-930-1
Sequence 1, Application US/08342930
Patent No. 5821084
GENERAL INFORMATION:
APPLICANT: OLIMSTED, ELIZABETH A.
APPLICANT: MAURO, LAURA J.
APPLICANT: DAVIS, ALAN R.
APPLICANT: DIXON, JACK E.
TITLE OF INVENTION: OSTEOBLAST-TESTICULAR PROTEIN TYROSINE
TITLE OF INVENTION: PHOSPHATASE
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/342,930
FILING DATE: 21-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KONSRI, ANTOINETTE F.
REGISTRATION NUMBER: 34,202
REFERENCE/DOCKET NUMBER: 20344-20975.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 813-5600
TELEFAX: (415) 494-0792

TELEX: 706141
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5455 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 205..5337
US-08-342-930-1

Query Match 11.4% Score 14; DB 1; Length 5455;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 19 gatgccttaacc 32
|||||
Db 3558 GATGCCTTACACC 3571

RESULT 25
US-09-112-450-3/c
Sequence 3, Application US/09112450
Patent No. 6120999
GENERAL INFORMATION:
APPLICANT: Abad, Antonio Jose C.
APPLICANT: Choi, Gil
APPLICANT: Calderone, Richard A.
TITLE OF INVENTION: Histidine Kinase Two-component in Candida albicans
FILE REFERENCE: PF393
CURRENT APPLICATION NUMBER: US/09/112,450
CURRENT FILING DATE: 1998-07-09
EARLIER APPLICATION NUMBER: 60/052,273
EARLIER FILING DATE: 1997-07-10
EARLIER APPLICATION NUMBER: 60/074,308
EARLIER FILING DATE: 1998-02-11
NUMBER OF SEQ ID NOS: 9
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 3
LENGTH: 8561
TYPE: DNA
ORGANISM: Homo sapiens
US-09-112-450-3

Query Match 11.4% Score 14; DB 3; Length 8561;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 aggtatgtaaatg 15
|||||
Db 1667 AGGTATGTAAATG 1654

RESULT 26
US-08-816-155B-1/c
Sequence 1, Application US/08816155B
Patent No. 5990091
GENERAL INFORMATION:
APPLICANT: TARTAGLIA, JAMES
APPLICANT: COX, WILLIAM I.
APPLICANT: GETTIG, RUSSELL R.
APPLICANT: MARTINEZ, HECTOR
APPLICANT: PROLETTI, ENZO
APPLICANT: PINCUS, STEVEN E.
TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND
TITLE OF INVENTION: METHODS OF MAKING AND USES THEREOF
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
STREET: 745 FIFTH AVENUE

CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10151
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/816,155B
FILING DATE: 12-MAR-1997
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: KOMALSKI, THOMAS J.
REGISTRATION NUMBER: 32,147
REFERENCE/DOCKET NUMBER: 454310-2990
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-588-0800
TELEFAX: 212-588-0500
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 10281 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-816-155B-1

Query Match 11.4%; Score 14; DB 2; Length 10281;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 attattccacaca 120
|||||
DB 1788 ATTATTCCACACA 1775

RESULT 27
US-09-079-587-1/c
Sequence 1, Application US/09079587
Patent No. 6130066
GENERAL INFORMATION:
APPLICANT: TARTAGLIA, JAMES
APPLICANT: COX, WILLIAM I.
APPLICANT: GETTIG, RUSSELL R.
APPLICANT: MARTINEZ, HECTOR
APPLICANT: MARLETTI, ENZO
APPLICANT: PINCUS, STEVEN E.
TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND
TITLE OF INVENTION: METHODS OF MAKING AND USES THEREOF
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
STREET: 745 FIFTH AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10151
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/079,587
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/816,155
FILING DATE: 12-MAR-1997
ATTORNEY/AGENT INFORMATION:

NAME: KOMALSKI, THOMAS J.
REGISTRATION NUMBER: 32,147
REFERENCE/DOCKET NUMBER: 454310-2990
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-588-0800
TELEFAX: 212-588-0500
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 10281 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-079-587-1

Query Match 11.4%; Score 14; DB 3; Length 10281;
Best Local Similarity 100.0%; Pred. No. 71;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 attattccacaca 120
|||||
DB 1788 ATTATTCCACACA 1775

RESULT 28
US-08-742-185-101
Sequence 101, Application US/08742185
Patent No. 6020476
GENERAL INFORMATION:
APPLICANT: Page, David C.
APPLICANT: Reijo, Renee
APPLICANT: Saxena, Richa
APPLICANT: Hawkins, Trevor
APPLICANT: Reeve, Mary Pat
TITLE OF INVENTION: DAZ: A GENE FAMILY ASSOCIATED WITH AZOOSPERMIA
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: US
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/742,185
FILING DATE: 30-OCT-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/690,734
FILING DATE: 31-JUL-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/310,429
FILING DATE: 22-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: WH194-07A2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 101:
SEQUENCE CHARACTERISTICS:
LENGTH: 43795 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)

US-08-742-185-101

Query Match 11.4%; Score 14; DB 3; Length 43795;
Best Local Similarity 100.0%; Pred. No. 72;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 100 attccattatt 113
|||||
Db 32551 ATCTCAATTATT 32564

RESULT 29

US-09-503-391-16/c
Sequence 16, Application US/09503391
Patent No. 6300091
GENERAL INFORMATION:
APPLICANT: Patton, David A.
APPLICANT: Ashby, Carl S.
APPLICANT: Thomas, Carla R.
APPLICANT: McEliver, John A.
APPLICANT: Budziszewski, Gregory J.
TITLE OF INVENTION: Herbicide Target Genes and Methods
FILE REFERENCE: PB/5-30852A
CURRENT FILING DATE: 2000-02-14
NUMBER OF SEQ ID NOS: 26
SOFTWARE: Patentln Ver. 2.1
SEQ ID NO 16
LENGTH: 30
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:
US-09-503-391-16

Query Match 10.6%; Score 13; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 111 attccacacaca 123
|||||
Db 16 ATTCACACACACA 4

RESULT 30

US-08-427-569-14/c
Sequence 14, Application US/08427569
Patent No. 6235465
GENERAL INFORMATION:
APPLICANT: Urdea, Michael S.
TITLE OF INVENTION: HTLV-1 PROBES FOR USE IN SOLUTION
NUMBER OF INVENTIONS: PHASE SANDWICH HYBRIDIZATION ASSAYS
NUMBER OF SEQUENCES: 55
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 755 Page Mill Road
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1018
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/427,569
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/130,150
FILING DATE:
APPLICATION NUMBER: 07/813,585
FILING DATE: 23-DEC-1991
ATTORNEY/AGENT INFORMATION:
NAME: Thomas E. Clotell
REGISTRATION NUMBER: 21,013
REFERENCE/DOCKET NUMBER: 22300-20238.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-813-5600
TELEFAX: 415-494-0792
TELEX: 706141

INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 33 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-427-569-14

Query Match 10.6%; Score 13; DB 4; Length 33;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 19 gatgcttaccac 31
|||||
Db 18 GATGCTTACAC 6

RESULT 31

US-08-311-760A-315
Sequence 315, Application US/08311760A
Patent No. 5599706
GENERAL INFORMATION:
APPLICANT: Scinichcomb, Dan T.
APPLICANT: McSwiggen, James
APPLICANT: Newton, Roger S.
APPLICANT: Ramharack, Randy
TITLE OF INVENTION: RIBOZYME TREATMENT OF DISEASES
TITLE OF INVENTION: OR CONDITIONS RELATED TO LEVELS OF
TITLE OF INVENTION: PLASMA LIPOPROTEIN (a) [LP(a)] BY
TITLE OF INVENTION: INHIBITING APOLOPROTEIN
NUMBER OF SEQUENCES: 392
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,760A
FILING DATE: September 23, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 208/155
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600

TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 315:
SEQUENCE CHARACTERISTICS:
LENGTH: 36 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-311-760A-315

Query Match 10.6%; Score 13; DB 1; Length 36;
Best Local Similarity 84.6%; Pred. No. 2.3e+02;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 27 acaacctgatga 39
|||||:|:|
Db 2 ACAACCCUGAUGA 14

RESULT 32

US-08-311-760A-316
Sequence 316, Application US/08311760A
Patent No. 5599706
GENERAL INFORMATION:
APPLICANT: Stinchcomb, Dan T.
APPLICANT: McSwigen, James
APPLICANT: Newton, Roger S.
TITLE OF INVENTION: RIBOZYME TREATMENT OF DISEASES
TITLE OF INVENTION: OR CONDITIONS RELATED TO LEVELS OF
TITLE OF INVENTION: PLASMA LIPOPROTEIN (a) [LP(a)] BY
TITLE OF INVENTION: INHIBITING APOLIPOPROTEIN
NUMBER OF SEQUENCES: 392
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/311,760A
FILING DATE: September 23, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 208/155
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 316:
SEQUENCE CHARACTERISTICS:
LENGTH: 36 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-311-760A-316

Query Match 10.6%; Score 13; DB 1; Length 36;

Best Local Similarity 84.6%; Pred. No. 2.3e+02;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 27 acaacctgatga 39
|||||:|:|
Db 2 ACAACCCUGAUGA 14

RESULT 33

US-08-774-310-315
Sequence 315, Application US/08774310
Patent No. 5877022
GENERAL INFORMATION:
APPLICANT: Stinchcomb, Daniel T.
APPLICANT: McSwigen, James
APPLICANT: Newton, Roger S.
TITLE OF INVENTION: RIBOZYME TREATMENT OF DISEASES
TITLE OF INVENTION: OR CONDITIONS RELATED TO LEVELS OF
TITLE OF INVENTION: PLASMA LIPOPROTEIN (a) [LP(a)] BY
TITLE OF INVENTION: INHIBITING APOLIPOPROTEIN
NUMBER OF SEQUENCES: 392
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
STREET: Suite 4700
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/774,310
FILING DATE: December 23, 1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/311,760
FILING DATE: September 23, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 223/229
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 315:
SEQUENCE CHARACTERISTICS:
LENGTH: 36 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-774-310-315

Query Match 10.6%; Score 13; DB 2; Length 36;
Best Local Similarity 84.6%; Pred. No. 2.3e+02;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 27 acaacctgatga 39
|||||:|:|
Db 2 ACAACCCUGAUGA 14

RESULT 34

US-08-774-310-316
Sequence 316, Application US/08774310
Patent No. 5877022

GENERAL INFORMATION:
APPLICANT: Slinchcomb, Daniel T.
APPLICANT: McSwiggen, James
APPLICANT: Newton, Roger S.
APPLICANT: Ramharack, Randy
TITLE OF INVENTION: RIBOZYME TREATMENT OF DISEASES
TITLE OF INVENTION: OR CONDITIONS RELATED TO LEVELS OF
TITLE OF INVENTION: PLASMA LIPOPROTEIN (a). [LP(a)] BY
TITLE OF INVENTION: INHIBITING APOLOPROTEIN
NUMBER OF SEQUENCES: 392
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/774,310
FILING DATE: December 23, 1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/311,760
FILING DATE: September 23, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Wardburg, Richard
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 223/229
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 316:
SEQUENCE CHARACTERISTICS:
LENGTH: 36 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-774-310-316

Query Match 10.6%; Score 13; DB 2; Length 36;
Best Local Similarity 84.6%; Pred. No. 2.3e+02;
Matches 11; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 27 acaaccctgatga 39
|||||:|:|
DB 2 ACAACCCUGAUGA 14

RESULT 35
US-08-207-226A-3/C
Sequence 3, Application US/08207226A
Patent No. 5580970
GENERAL INFORMATION:
APPLICANT: Hendricks, David
APPLICANT: Rigby, Susan
APPLICANT: Parodos, Kyriaki
TITLE OF INVENTION: DETECTION OF HPV TRANSCRIPTS
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amoco Corporation
STREET: 55 Shuman Boulevard, Suite 600
CITY: Naperville
STATE: Illinois
COUNTRY: US

ZIP: 60563
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/207,226A
FILING DATE: 04-MAR-1994
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/622,742
FILING DATE: 05-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Galloway, No. 5580970val B.
REFERENCE/DOCKET NUMBER: GTR-8903AF
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 39 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-207-226A-3

Query Match 10.6%; Score 13; DB 1; Length 39;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ttgtaattgtgag 19
|||||
DB 30 TTGTAATGTGAG 18

RESULT 36
US-08-425-684-133/C
Sequence 133, Application US/08425684
Patent No. 5834252
GENERAL INFORMATION:
APPLICANT: STEMMER PH.D., WILLEM P.C.
TITLE OF INVENTION: END-COMPLEMENTARY POLYMERASE REACTION
NUMBER OF SEQUENCES: 136
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW
STREET: ONE MARKET PLAZA, STEWART STREET TOWER, 20TH FLOOR
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: U.S.A.
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425,684
FILING DATE: 18-APR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: DUNN ESQ., TRACY J.
REGISTRATION NUMBER: 34,587
REFERENCE/DOCKET NUMBER: 165284-015400US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
INFORMATION FOR SEQ ID NO: 133:
SEQUENCE CHARACTERISTICS:
LENGTH: 40 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
US-08-425-684-133

Query Match 10.6%; Score 13; DB 2; Length 40;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 attccacacaca 123
Db 21 ATTCCACACACACA 9

RESULT 37

US-08-675-502-133/c
; Sequence 133, Application US/08675502
; Patent No. 5928905
; GENERAL INFORMATION:
; APPLICANT: STEMMER PH.D., WILLEM P.C.
; APPLICANT: LIPSHUTZ, ROBERT J.
; TITLE OF INVENTION: END-COMPLEMENTARY POLYMERASE REACTION
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND AND TOWNSEND AND CREW
; STREET: TWO EMBARCADERO CENTER, 8TH FLOOR
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: U.S.A.
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/675,502
; FILING DATE: 03-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/425,684
; FILING DATE: 18-APR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/05480
; FILING DATE: 18-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MURPHY, ESO., MATTHEW B.
; REGISTRATION NUMBER: 39,787
; REFERENCE/DOCKET NUMBER: 16528J-015410US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 326-2400
; TELEFAX: (415) 326-2422
; INFORMATION FOR SEQ ID NO: 133:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 40 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; US-08-675-502-133

Query Match 10.6%; Score 13; DB 2; Length 40;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 attccacacaca 123
Db 21 ATTCCACACACACA 9

RESULT 38

US-09-129-740-10/c
; Sequence 10, Application US/09129740A
; Patent No. 6067246

; GENERAL INFORMATION:
; APPLICANT: HELLER, MICHAEL J.
; APPLICANT: TU, EUGENE
; TITLE OF INVENTION: DNA OPTICAL STORAGE
; FILE REFERENCE: DAVID B. MURPHY - Heller 236/119
; CURRENT APPLICATION NUMBER: US/09/129,740A
; CURRENT FILING DATE: 1998-08-05
; EARLIER APPLICATION NUMBER: 08/906,569
; EARLIER FILING DATE: 1997-08-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 10
; LENGTH: 42
; TYPE: DNA
; ORGANISM: synthetic construct
; US-09-129-740-10

Query Match 10.6%; Score 13; DB 3; Length 42;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 attccacacaca 123
Db 30 ATTCCACACACACA 18

RESULT 39

US-08-821-559A-17/c
; Sequence 17, Application US/08821559A
; Patent No. 5846774
; GENERAL INFORMATION:
; APPLICANT: XIA, YUANNAN
; TITLE OF INVENTION: CHLORELLA VIRUS PROMOTERS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 3100 No. 5846774west Center, 90 South Seventh st
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/821,559A
; FILING DATE: 21-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Kettelberger, Denise M
; REGISTRATION NUMBER: 33,924
; REFERENCE/DOCKET NUMBER: 8648.63-US-01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612/371-5268
; TELEFAX: 612/332-9081
; TELEX:
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-821-559A-17

Query Match 10.6%; Score 13; DB 2; Length 50;

Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 attccacacaca 123
|||||
DB 49 ATTCCACACACA 37

RESULT 40

PCT-US95-14418-50
Sequence 50, Application PC/TUS9514418
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: DNA Encoding a Thermostable DNA Polymerase Enzyme
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/14418
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Gass, David A.
REGISTRATION NUMBER: 38,153
REFERENCE/DOCKET NUMBER: 28003/32330
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 70 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US95-14418-50

Query Match 10.6%; Score 13; DB 5; Length 70;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 attccacacaca 123
|||||
DB 57 ATTCCACACACA 69

RESULT 41

PCT-US95-15327-50
Sequence 50, Application PC/TUS9515327
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Biologically Active Fragments of
TITLE OF INVENTION: Therms Flavus DNA Polymerase
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 6300 Sears Tower, 233 South Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/15327
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Gass, David A.
REGISTRATION NUMBER: 38,153
REFERENCE/DOCKET NUMBER: 28003/31716
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312/474-6300
TELEFAX: 312/474-0448
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 70 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US95-15327-50

Query Match 10.6%; Score 13; DB 5; Length 70;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 attccacacaca 123
|||||
DB 57 ATTCCACACACA 69

RESULT 42
US-08-569-147-69
Sequence 69, Application US/08569147
Patent No. 6180377
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: HUMANISED ANTIBODIES
NUMBER OF SEQUENCES: 95
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz &
ADDRESSEE: No. 6180377ris, LLP
STREET: One Liberty Place - 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: U.S.A.
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/569,147
FILING DATE: 25-March-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Trujillo, Doreen Yalko
REGISTRATION NUMBER: 35,719
REFERENCE/DOCKET NUMBER: CAP-0047
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 71 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

MOLECULE TYPE: CDNA
US-08-569-147-69

Query Match 10.6%; Score 13; DB 4; Length 71;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 agatgccttaca 30
DB 28 AGATGCCTTACAA 40

RESULT 43

US-09-235-246-21/C
Sequence 21, Application US/09235246A
Patent No. 6048719

GENERAL INFORMATION:

APPLICANT: Kong, Huimin

APPLICANT: Higgins, Lauren S.

APPLICANT: Dalton, Michael A.

TITLE OF INVENTION: Method For Cloning And Producing The DraIII Restriction

TITLE OF INVENTION: Endonuclease

FILE REFERENCE: DraIII

CURRENT APPLICATION NUMBER: US/09/235,246A

CURRENT FILING DATE: 1999-01-22

NUMBER OF SEQ ID NOS: 21

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 21

LENGTH: 95

TYPE: DNA

ORGANISM: Deinococcus radiophilus

US-09-235-246-21

Query Match 10.6%; Score 13; DB 3; Length 95;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 111 attccacacaca 123
DB 57 ATTCCACACACA 45

RESULT 44

US-08-253-877C-12/C

Sequence 12, Application US/08253877C
Patent No. 5773001

GENERAL INFORMATION:

APPLICANT: Hamann, Phillip R.

APPLICANT: Hinman, Lois

APPLICANT: Hollander, Irwin

APPLICANT: Holcomb, Ryan

APPLICANT: Hallett, William

APPLICANT: Tsou, Hwei-Ru

APPLICANT: Weiss, Martin J.

TITLE OF INVENTION: Conjugates of Methyltrithio Antitumor

TITLE OF INVENTION: Agents and Intermediates for Their Synthesis

NUMBER OF SEQUENCES: 73

CORRESPONDENCE ADDRESS:

ADDRESSEE: American Cyanamid Company

STREET: One Cyanamid Plaza

CITY: Wayne

STATE: New Jersey

COUNTRY: U.S.A.

ZIP: 07470-8426

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/253,877C

FILING DATE: 03-JUN-1994

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Barnhard, Elizabeth M.

REGISTRATION NUMBER: 31,088

REFERENCE/DOCKET NUMBER: 32,368

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-831-3246

TELEFAX: 201-831-3305

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 96 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-253-877C-12

Query Match 10.6%; Score 13; DB 1; Length 96;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 18 agatgccttaca 30
DB 78 AGATGCCTTACAA 66

RESULT 45

US-08-452-164A-12/C

Sequence 12, Application US/08452164A
Patent No. 5877296

GENERAL INFORMATION:

APPLICANT: Hamann, Phillip R.

APPLICANT: Hinman, Lois

APPLICANT: Hollander, Irwin

APPLICANT: Holcomb, Ryan

APPLICANT: Hallett, William

APPLICANT: Tsou, Hwei-Ru

APPLICANT: Weiss, Martin J.

TITLE OF INVENTION: Conjugates of Methyltrithio Antitumor

TITLE OF INVENTION: Agents and Intermediates for Their Synthesis

NUMBER OF SEQUENCES: 73

CORRESPONDENCE ADDRESS:

ADDRESSEE: American Home Products Corporation

STREET: One Campus Drive

CITY: Parsippany

STATE: New Jersey

COUNTRY: U.S.A.

ZIP: 07054

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/452,164A

FILING DATE: 26-MAY-1995

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Barnhard, Elizabeth M.

REGISTRATION NUMBER: 31,088

REFERENCE/DOCKET NUMBER: 32,368-04

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-683-2158

TELEFAX: 201-683-4117

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 96 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-452-164A-12

Query Match 10.6%; Score 13; DB 2; Length 96;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 18 agatgccttaca 30
|||||
Db 78 AGATGCCTTACAA 66

Search completed: March 19, 2002, 11:04:27
Job time: 3856 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2002, 09:31:36 ; Search time 1345.18 Seconds
(without alignments)
1508.462 Million cell updates/sec

Title: US-09-684-016-48411

Perfect score: 123
Sequence: 1 caggtattgtaaatgtaga.....tcaattatccacacaca 123

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:
1: gb_da:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_hlgo_hum:*
31: em_hlgo_inv:*
32: em_hlgo_rod:*
33: em_hlgo_hum:*
34: em_hlgo_inv:*
35: em_hlgo_rod:*
36: em_hlgo_other:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	54.4	44.2	89035	8	ATT22F8	AL050351 Arabidops
C 2	54.4	44.2	100469	8	ATF19H22	AL035679 Arabidops
C 3	54.4	44.2	198987	8	ATCHRIV90	AL161594 Arabidops
C 4	53.4	43.4	103270	8	ATFLN20	AL022140 Arabidops
C 5	53.4	43.4	196247	8	ATCHRIV56	AL161556 Arabidops
C 6	52.4	42.6	119807	8	AP003753	AP003753 Oryza sat
C 7	47.4	38.2	1062	8	AF277453	AF277453 Arabidops
C 8	45.4	36.9	1058	8	HVES43	X77575 H. vulgare (
C 9	41.8	34.0	367	8	AB011445	AB011445 Trilicium
C 10	34.4	28.0	44723	6	AX059527	AX059527 Sequence
C 11	34.4	28.0	80446	8	T27D20	AE076274 Arabidops
C 12	34.4	28.0	198022	8	ATCHRIV11	AL161499 Arabidops
C 13	33.8	27.5	154195	9	AL161773	AL161773 Human DNA
C 14	33.8	27.5	166266	2	AP001146	AP001146 Homo sapi
C 15	33.8	27.5	185404	2	AC025409	AC025409 Homo sapi
C 16	31.6	25.7	265815	2	AC027308	AC027308 Homo sapi
C 17	31.4	25.5	799	5	GGGP23	X60795 Chicken mRN
C 18	31.2	25.4	23366	1	AE002311	AE002311 Chlamydia
C 19	31	25.2	30494	3	CEK10D6	Z74040 Caenorhabdl
C 20	30	24.4	83145	8	AC002339	AC002339 Arabidops
C 21	29.8	24.2	68575	2	AC027346	AC027346 Homo sapi
C 22	29.8	24.2	161818	2	AP003409	AP003409 Oryza sat
C 23	29.8	24.2	170020	8	AP003256	AP003256 Oryza sat
C 24	29.6	24.1	63904	2	AC017284	AC017284 Drosophila
C 25	29.6	24.1	80065	2	AC016271	AC016271 Homo sapi
C 26	29.6	24.1	94797	3	AC005333	AC005333 Drosophila
C 27	29.6	24.1	166441	3	AE003611	AE003611 Drosophila
C 28	29.6	24.1	259973	3	AE003611	AE003611 Drosophila
C 29	29.4	23.9	155172	2	AC021110	AC021110 Homo sapi
C 30	29.4	23.9	168656	9	AL359922	AL359922 Human DNA
C 31	29.4	23.9	174815	2	AP001350	AP001350 Homo sapi
C 32	29.4	23.9	178056	2	AC091589	AC091589 Homo sapi
C 33	29.4	23.9	195491	2	AL449423	AL449423 Homo sapi
C 34	29.2	23.7	120515	9	AL136230	AL136230 Human DNA
C 35	29.2	23.7	159073	2	AC013621	AC013621 Homo sapi
C 36	29.2	23.7	159150	2	AC022863	AC022863 Homo sapi
C 37	29.2	23.7	170022	2	AC021559	AC021559 Homo sapi
C 38	29.2	23.7	181127	2	AC068203	AC068203 Homo sapi
C 39	29.2	23.7	181542	2	AC023187	AC023187 Homo sapi
C 40	28.8	23.4	1086	8	VAPPCGENE	X91634 V. apyalla m
C 41	28.8	23.4	14039	2	AC018023	AC018023 Drosophila
C 42	28.8	23.4	17078	1	AE001152	AE001152 Borrelia
C 43	28.8	23.4	49759	3	AE002717	AE002717 Drosophila
C 44	28.8	23.4	74534	3	AC004735	AC004735 Drosophila
C 45	28.8	23.4	85195	9	AL136093	AL136093 Human DNA

ALIGNMENTS

RESULT 1
ATT22F8/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
JOURNAL
REFERENCE
AUTHORS
TITLE

89035 bp DNA
Arabidopsis thaliana DNA chromosome 4, BAC clone T22F8 (ESSA project).
AL050351.1 GI:4914422
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidops.
1 (bases 1 to 89035)
Bevan, M., Zimmermann, W., Gruenisen, A., Wambutt, R., Bancroft, I., Mewes, H.W., Meyer, K.F.X. and Schellier, C.
Unpublished
2 (bases 1 to 89035)
EU Arabidopsis sequencing, project.
Direct Submission

27-MAY-1999
PLN
27-MAY-1999
PLN
27-MAY-1999
PLN

JOURNAL Submitted (27-MAR-1999) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: schueller@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de, Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK, E-mail: michael.bevan@bbsrc.ac.uk

COMMENT Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/chal/>.

FEATURES
source location/Qualifiers
1. .89035
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
1. .7016
/note="overlap to BAC F19H22: please refer to EMBL acc. AL035679 for analysis and annotation"
complement(5175..7811)
/gene="T22F8.10"
5175..7811
/gene="T22F8.10"
complement(5175..7811)
/note="T22F8.10"
/note="similarity to receptor-like protein kinase, Catharanthus roseus, Z73295"
/codon_start=1
/product="putative receptor-like protein kinase"
/protein_id="CAB43626.1"
/db_xref="GI:4914423"
/translation="MEIRKKRNIFVLYIDSSKRSALLAILFLSPSASAVAA AVGATGKRPADDLIDGSSSSKTPDGRVFKSQETIYIEAKEDIQVSPSDKY ASPIYLARIEREATYKFLTRPGMAWRHLFLAPDKEPDLQATFSLVEKYL HNFKITSNNNDQSAVQKELYVNMTDAQFALFRPKSSAAFINALEVSADEILSD SGTALEPVIQSGSDYAYOSVRYNAGCPLIMPNDIGRTIMDEKLEKDENLAKD VKTPSAIKYPEVTPLIAPQTVATAVEMANSILIDNPNSMNPSPNSNYLIRL HECDIVSKSLNDIENYINNGKTAISGDLSTVAGNLAAPYKDIVNATILMGPELOY QICPMGEDTGTKNAILNGVEVLMKNSVNSLDGEGVIGRTTGKRGHMAVATAGVMM FGAFILGAMAVYKMKRKPQDMQKRNFSFSSWLLPIHAGDSTFWTSKGSOKSFYNSL GLGRYFSLSELOEATKNEASQIIGVGSGFVNYIGTLDGRTVAVKRCNPQSEQCITE FQETIOMLSLRHRLIVSLIGCDENSEMILYVEPMSPGPDHLYGNLAFLTKOR LEICIGSARGHLVHTGAOGIIHRDVKSTNILLDEALVAVADRGLSKDAFFGONHV STAVKSGFYLDPEYFRROQLFDKSDVSEFVILAEALCARPAINPOLPREVYLAIEV AMQKRRGLLEKTIIDPLAGTINPSSKRFAPAAKCELDYGVDRPTMGDVLNMLEYA LQIQEAFYQGAETENAKRPVYTPGVSVSDPSPTSPSTVTTNEATVPVPAKVEENS GTAVDEHSGTAMFTGFANLNGR"
complement(5175..7811)
/gene="T22F8.10"
/number=1
8270..10413
/gene="T22F8.20"
complement(8270..8340)
/gene="T22F8.20"
/number=1
complement(join(8270..8340,8460..8571,8701..8790,9059..9125,9296..9420,9594..9643,9732..9826,10010..10088,10242..10413))
/gene="T22F8.20"
complement(join(8270..8340,8460..8571,8701..8790,9059..9125,9296..9420,9594..9643,9732..9826,10010..10088,10242..10413))
/gene="T22F8.20"
/note="similarity to Mono-phosphatase, Streptomyces anulatus, X92429
contains inositol monophosphatase family signatures [WVLDPIDGTRKSFRT][VLDPIDGTRKSFRT]"
/codon_start=1
/product="inositol monophosphatase-like protein"
/protein_id="CAB43627.1"
/db_xref="GI:4914424"
/translation="MASNRPNISNESPSSELTDELDFPAVGNALADASGEVIRKY FRKKFDIVDKDMSPTIADQMAEAMWSIIFONPSIAITGEEKGMCKGKESADVYV VLDPIDGTRKSFRTTKPVFGTILAILYKRPILGLIDPILKERWIGMGRKRLNGED

ISTRSCPULSQAYLYTTSPLHPSERAEAKYSRVDKVKVPIYGCDCYAYALLASGFD IVALSEGLKPYDFLALVPVIEAGGTTIDWTKRFLMEASSSAVATSFNVAAAGDSDIH IVALSEGLKPYDFLALVPVIEAGGTTIDWTKRFLMEASSSAVATSFNVAAAGDSDIH

complement(8341..8459)
/number=1
complement(8460..8571)
/gene="T22F8.20"
/number=2
complement(8572..8700)
/number=2
complement(8701..8790)
/gene="T22F8.20"
/number=3
complement(8791..9058)
/number=3
complement(9059..9125)
/gene="T22F8.20"
/number=4
complement(9126..9295)
/number=4
complement(9296..9420)
/gene="T22F8.20"
/number=5
complement(9421..9593)
/number=5
complement(9594..9643)
/gene="T22F8.20"
/number=6
complement(9644..9731)
/number=6
complement(9732..9826)
/gene="T22F8.20"
/number=7
complement(9827..10009)
/number=7
complement(10010..10088)
/gene="T22F8.20"
/number=8
complement(10089..10241)
/number=8
complement(10242..10413)
/gene="T22F8.20"
/number=9
complement(join(10875..11060,11267..11536))
/gene="T22F8.30"
10875..11536
/gene="T22F8.30"
complement(join(10875..11060,11267..11536))
/gene="T22F8.30"
complement(join(10875..11060,11267..11536))
/note="similarity to dehydrin, glycine max, AF004807"
/codon_start=1
/product="putative protein"
/protein_id="CAB43628.1"
/db_xref="GI:4914425"
/translation="MADLDKDERGSPYITLDAHGBPAQMDLDFGNAMHLTGATVPHL KESSTYGHPIPTAYVTITNTPFHNAQPIYSVSDPQLQDDLDLHFGTSSLEEGBEVGKRT NITDETAKSLGVDPASATVYGGSSGVHEKKGFKIKIKELSGHNDL"
complement(10875..11060)
/gene="T22F8.30"
/number=1
complement(11061..11266)
/number=1
complement(11267..11536)
/gene="T22F8.30"
/number=2
complement(12329..13285)
/gene="T22F8.40"
/number=1
12329..13974
/gene="T22F8.40"
complement(join(12329..13285,13369..13415,13689..13974))
/gene="T22F8.40"
complement(join(12329..13285,13369..13415,13689..13974))

Query Match	Best Local Similarity	Score	DB	Length
Matches 64; Conservative	80.0%;	44.28;	DB 8;	89035;
	0; Mismatches 16;	Pred. No. 2.9e-08;		Indels 0; Gaps 0;
OY 1	caggtatgttaaatgtagatgccttaacacccgatgacccaatgagcgcaatgcaagg 60			
Db 2099	CAGGTCGCAAGGTGTGAGATGCCGTATACCCAGCATGACTGATGATGCAATGCGAGGA 2040			
OY 61	ctgcactgactgctgaattt 80			
Db 2039	GTCGTCGATGATGCTGATGCT 2020			
RESULT 2	ATP19H22	100469 bp	DNA	PLN 27-AUG-1999
LOCUS	Arabidopsis thaliana DNA chromosome 4, BAC clone F19H22 (ESSA project).			
DEFINITION	Arabidopsis thaliana DNA chromosome 4, BAC clone F19H22 (ESSA project).			
ACCESSION	AL035679			
VERSION	AL035679.1	GI:4539309		
KEYWORDS				
SOURCE	thale cress.			
ORGANISM	Arabidopsis thaliana			
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eustosids II; Brassicales; Brassicaceae; Arabidopsis. 1 (bases 1 to 100469)			
AUTHORS	Bayan, M., Murphy, G., Ridley, P., Hudson, S., Bancroft, I., Mewes, H.W., et al.			
JOURNAL	Mayer, K.F.X. and Schueler, C. .			
REFERENCE	Unpublished			
AUTHORS	2 (bases 1 to 100469)			
TITLE	EU Arabidopsis sequencing project.			
JOURNAL	Submitted (11-MAR-1999) MRPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitze 10a, D-82152 Martinsried, FRG. E-mail: schueler@mpg.de, mayer@mpg.de, biochem.mpg.de			

FEATURES	source
COMMENT	<p>Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NN4 7UJ Norwich, UK, E-mail: michael.bevan@brc.ac.uk Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosome 4 can be viewed at: http://webery.mips.biochem.mpg.de/proj/thal/.</p>
FEATURES	<p>location/Qualifiers</p> <p>1. 100469</p>
source	<p>/organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4"</p>
misc_feature	<p>1. 17740 /note="Overlap to BAC T9A14, please refer to this entry for analysis and annotation" complement(join(20670..21072,21414..21540,21709..22405,22506..22616)) /gene="F19H22.10" 20670..22616 /gene="F19H22.10" complement(20670..21072) /gene="F19H22.10" number=1 complement(join(20670..21072,21414..21540,21709..22405,22506..22616)) /gene="F19H22.10" /note="similarity to T32F12.7 Arabidopsis thaliana chromosome II BAC T32F12, PID:g3608133" /codon_start=1 /product="putative protein" /protein_id="CAB38811.1" /db_xref="GI:4539310" /translation="MPDINFSIAVESKRMEGGOYENGRYKPDYKYKQTQSVNPKPEQHVALVNNKIIISLIERDAVKEKNEAVATKEALSRDEALRQDKALSERDNALIMETESALVALRYRENNLNYILSCARGGSGRFTIEESLIPNPSISTIPPAANTPRPKRKESGOGKRMGDLNRPVASPGKSKDMDSNDVIATPDGMTMPVMCCTGAROCYKNGNGMOSSCCTTILSEYPLDPMPKRRSRVGGKRMSSSVSRLLSRADLTRKSRKDEKRLVIVFTVDSVKRQISSSTVLWFCCEFTATFFSGIAFAEFTLNKHNHSGWMIIVYALSLIILPLIMGLGIWIFESSHSQMVQAFPMVEEDCVGSARSGNKKETKDNKCESLAIYVNSADQKCKMEKTFENKRTAKLTVSFPLHSQVRSCTR"</p>
intron	<p>/number=1 complement(21414..21540) /gene="F19H22.10"</p>
exon	<p>/number=2 complement(21541..21708) /number=2 complement(21709..22405) /gene="F19H22.10"</p>
intron	<p>/number=3 complement(22406..22505) /number=3 complement(22506..22616) /gene="F19H22.10"</p>
exon	<p>/number=4 23475..23550 /gene="F19H22.20"</p>
exon	<p>/number=1 23475..24998 /gene="F19H22.20" join(23475..23550,24243..24528,24866..24998) /gene="F19H22.20"</p>
gene	<p>/EC number="3.6.1.35" /note="one AA difference to PIR2:S60132 contains EST gb:T45164, R90510, T46102, F14129"</p>
CDS	<p>/codon_start=1 /product="H+-transporting ATPase 16k chain p2, vacuolar" /protein_id="CAB38812.1" /db_xref="GI:4539311" /translation="MSFSGDDETAPFGPLGAALVFCMGAAVGTAKSGVAVASGKVMPELVMSKIVPVNAGVIGTGLIIVAIISGINKAKSYLLDFQYAHLSGIACGLAELSGMAIGIVDAGVRANAOQPKLVEGMILLIFAEALALGLIVGLITLSSRQGLSRAE"</p>

```

intron      23551..24242
            /gene="F19H22.20"
            /number=1
exon        24243..24528
            /gene="F19H22.20"
            /number=2
intron      24529..24865
            /gene="F19H22.20"
            /number=2
exon        24866..24998
            /gene="F19H22.20"
            /number=3
gene        26052..27761
            /gene="F19H22.30"
            /note="ESTs GB:227052, 229899, H76778, T76313, H76773,
            T76734 are in 3' untranslated region
            EST T42894 covers intron in 5' untranslated region
            strong similarity to ubiquitin fusion-degradation protein
            - Mus musculus, PID:q1654348
            contains EST gb:AA395284, T42894"
            /codon_start=1
            /product="putative ubiquitin-dependent proteolytic
            protein"
            /protein_id="CAB38813.1"
            /db_xref="GI:4539312"
            /translation="MFYDGYAHGTTPEOTYRCYPSFIDKPOIESGDKIIMPSPALD
            RLASLIDYPMLELRNASTDSFSHGVLFEIAEGVITYIPYMMONILLQSGDVRV
            RNVLPKCTYKLOPHTPLDIANPKAILETALRNYSCLTYGDSIMPYNNKKTFID
            IVEKPSGISITIEDCEVDPAFLDIKEPERKPAVAPAKGAKAEVDVAAEKPFN
            PFTSGRRLDGRPLSEPOPGAAANSQSHPYAVASSSSGSEKATQONRCKLVEGSNVE
            RSTETTKVGAGKRRKOEAEKAKFOAFSGKKYSLNG"
exon        26052..26054
            /gene="F19H22.30"
            /number=1
intron      26055..26137
            /gene="F19H22.30"
            /number=1
exon        26138..26215
            /gene="F19H22.30"
            /number=2
intron      26216..26339
            /gene="F19H22.30"
            /number=2
exon        26340..26364
            /gene="F19H22.30"
            /number=3
intron      26365..26464
            /gene="F19H22.30"
            /number=3
exon        26465..26497
            /gene="F19H22.30"
            /number=4
intron      26498..26588
            /gene="F19H22.30"
            /number=4
exon        26589..26710
            /gene="F19H22.30"
            /number=5
intron      26711..26792
            /gene="F19H22.30"
            /number=5
exon        26793..26923
            /gene="F19H22.30"
            /number=6
intron      26924..27037
            /gene="F19H22.30"
            /number=6
exon        27038..27258
            /gene="F19H22.30"

```

```

intron      /number=7
            27259..27333
            /gene="F19H22.30"
            /number=7
exon        27334..27572
            /gene="F19H22.30"
            /number=8
intron      27573..27665
            /gene="F19H22.30"
            /number=8
exon        27666..27761
            /gene="F19H22.30"
            /number=9
repeat_region 27950..27971
            /note="22bp T tandem repeat"
exon        28993..30105
            /gene="F19H22.40"
            /number=1
gene        28993..30105
            /gene="F19H22.40"
            /note="similarity to more than 30 predicted proteins,
            Arabidopsis
            contains EST gb:AB015106"
            /codon_start=1
            /product="putative protein"
            /protein_id="CAB38814.1"
            /db_xref="GI:4539313"
            /translation="MSPPIIRAPKKOPPEPCLISLPEEIVVDIYARVRCYPTL
            SVSRFRFSIAPSEIYKRRPFCGTECLYLAISKOTSDIHMTLCRPNQSGS
            TTASHRLVHPTLPMPMHSGSYGIGSNITVKGFCMKRTTSVSLDCTRHAQTL
            PMPRAVAFVPELIDRKIRIVIGSDTLSPKSPSRIMMVYDTDTKMQULNARDWEA
            QY 1 caggatctgtaaatgtgagatgcctacaacccatgacctaatgagcgcaatgcgaagg 60
            ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
            Db 95552 CAGGTTCTGCAAGTGTGAGATGCCGTATACCAGATGACTTGATGATCGCAGGA 95493
            QY 61 ctgcactgactgtaagttt 80
            || || ||||| ||||| ||
            Db 95492 GTTCTGTGAGTGTGATGCT 95473

RESULT 3
LOCUS ATCHRIV90 198987 bp DNA PLN 16-MAR-2000
DEFINITION Arabidopsis thaliana DNA chromosome 4, contig fragment No. 90.
ACCESSION AL161594
VERSION AL161594.2 GI:7270858
KEYWORDS
SOURCE
ORGANISM
    thale cress.
    Arabidopsis thaliana
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
    Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
    1. (bases 1133 to 1779)
    Robben,J., Braeken,M., Glynnoprez,B., Volckaert,G., Mewes,H.W.,
    Lemcke,K. and Mayer,K.F.X.
    JOURNAL Unpublished
    2. (bases 131465 to 198987)
    Zimmermann,W., Grueniseisen,A., Wambutt,R., Kalicki,J., Wohlmann,P.,
    Smith,A., Mewes,H.W., Lemcke,K. and Mayer,K.F.X.
    JOURNAL Unpublished
    3. (bases 1 to 55749; 38011 to 138479)
    Murphy,G., Ridley,P., Hudson,S., Mewes,H.W., Lemcke,K. and
    Mayer,K.F.X.
    JOURNAL Unpublished
    4. (bases 1 to 198987)
    EU Arabidopsis sequencing project.

```

TITLE	JOURNAL
Direct Submission Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de , mayer@mips.biochem.mpg.de , Project Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge Laboratory, John Innes Centre, Colney Lane, NR4 7UD Norwich, UK, E-mail: michael.bevan@bsrc.ac.uk Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thel/ this frgmet has an overlap with ATCHRIY91 at the 3' end and an overlap with ATCHRIY91 at the 5' end.	
COMMENT	
FEATURES	
source	/organism="Arabidopsis thaliana" /variety="Columbia" /db_xref="taxon:3702" /chromosome="4" complement(join(5956..6064,6141..6226,6420..6539,6642..6843,7032..7189,7275..7334,7417..7494,7670..7768,7851..8019,8162..8319,8693..8828,9002..9122,9484..9538,9581..9688,9729..9755)) /gene="AtAg38760" 5956..9755 /gene="AtAg38760" complement(join(5956..6064,6141..6226,6420..6539,6642..6843,7032..7189,7275..7334,7417..7494,7670..7768,7851..8019,8162..8319,8693..8828,9002..9122,9484..9538,9581..9688,9729..9755)) /gene="AtAg38760" /note="similarity to KIAA0169 gene, Homo sapiens, D79991" /codon_start=1 /product="putative protein" /protein_id="CA80539.1" /db_xref="GI:7270839"
CDS	/translation="MDVEKTDIMNDEIOSYLVERSEOEYGTTSVAOELTQEFIDMSFHLINOCFFVWWSCEAFLLTLSPFNVLIEDIASVLHTAPVTKYRKASHTCICSYVARRESSIKKEAVKLISDGLERROSSVEDLSSCPKNMGITSGSNFSFKLVASVOHSAACROVOIOLMLILFTLEDMENLOWHDGVPRSCVPSTVDONEMA TTSSLNTSEVNAGPLVANAFCLDISLPKRSPFLMIDIVSVYHOAEFASIS YFEILLQSRLNDPFGPISGHSVVRTISAFASETNLQLEDGLELLIDLISKY QGEESLCDFMRKSFPVDGPFRCLLFDESERPERSAFIRLSLSGSMPACVNA PLDKSVGYSTLEFDITSPPADASQLVETSRPLHPGLEGLVIPSTRGRILRVISEN TVLYRREYSLSGITLIILTRANKXIGNNRPAFVTELRMTVFENKAVCPSLINISH FTVYGESYVNGKMSEDVRVIDICTNSRSLTFGDSGAAYMAIDIILAKLLRW"
exon	complement(5956..6064) /gene="AtAg38760"
intron	/number=1 complement(6065..6140)
exon	/number=1 complement(6141..6226) /gene="AtAg38760"
intron	/number=2 complement(6227..6419)
exon	/number=2 complement(6420..6539) /gene="AtAg38760"
intron	/number=3 complement(6540..6641)
exon	/number=3 complement(6642..6843) /gene="AtAg38760"
intron	/number=4 complement(6844..7031)
exon	/number=4 complement(7032..7189) /gene="AtAg38760"
intron	/number=5 complement(7190..7274)
exon	/number=5 complement(7275..7334) /gene="AtAg38760"
intron	/number=6 complement(7335..7416) /number=6 complement(7417..7494) /gene="AtAg38760"
exon	/number=7 complement(7495..7669) /number=7 complement(7670..7768) /gene="AtAg38760"
intron	/number=8 complement(7769..7850)
exon	/number=8 complement(7851..8019) /gene="AtAg38760"
intron	/number=9 complement(8020..8161)
exon	/number=9 complement(8162..8319) /gene="AtAg38760"
intron	/number=10 complement(8320..8692)
exon	/number=10 complement(8693..8828) /gene="AtAg38760"
intron	/number=11 complement(8829..9001)
exon	/number=11 complement(9002..9122) /gene="AtAg38760"
intron	/number=12 complement(9123..9483)
exon	/number=12 complement(9484..9538) /gene="AtAg38760"
intron	/number=13 complement(9539..9580)
exon	/number=13 complement(9581..9688) /gene="AtAg38760"
intron	/number=14 complement(9689..9728)
exon	/number=14 complement(9729..9755) /gene="AtAg38760"
intron	/number=15 complement(9756..11163)
gene	/gene="AtAg38770" complement(join(11163..12361,12455..12602))) /gene="AtAg38770"
CDS	/note="similarity to proline-rich protein, Solanum tuberosum, AJ000997 contains EST gb:N66646, Z25647, T42523, AI995782.1, N65249, AA651424, Z33774, H37643, T42522, R65274, A1100294" /codon_start=1 /product="extensin-like protein" /protein_id="CA80540.1" /db_xref="GI:7270860"
intron	/translation="WKILPEPGSVPCILLVSLASTLSLARVEEVGYAESKIKT PAAFSGRLATIDCKVNGKHFWTKSGSDGDIGLGNOKLTPEDIVSDNGLAKECYAOQH SAAGTFCRAHDDTESTKITVFLSKSGDKHILGLIOKLKPSICEVSKEFWPKALPFEPKK GFDNPRLRPRLPLEPRLKRCPRKSPVPVPEVPEPKKEIPRRPVVDPDRPPV KKDEVRRPVVKRRKVLRPRPTIPKRCRPRKIEHRPVRVYKRPKITEKRRPVVYK RP RKPKIEHRPVRVYNLKDKCRPKKADPPRRVYNKPRTKRCRPRKCVDDRPRVYNK PPPKVIIPRKIEHRPVRVYKRPKIEHRPTVLRPYKRCRPRPRVLYKRPVVIKRP PCRPVRYVYKRPVVIKRPCRPRPOLRPLRFKRPFLRPKYTHNKFGKWPLRPBP"
exon	complement(11163..12361) /gene="AtAg38770"
intron	/number=1 complement(12362..12454) /number=1 complement(12455..12602) /gene="AtAg38770"

[illegible]

33259.. 32607)"/
/gene="AT4g21990"
/note="complements functionally an Escherichia coli mutant defective in PAPS reductase activity
Contains Prokaryotic membrane lipoprotein lipid attachment site AA276-286
contains EST gb:234086, Z35366"
/codon_start=1
/product="PRH26 protein"
/protein_id="CAI8102.1"
/db_xref="GI:2961344"
/db_xref="SPTREMBL:O48887"
/translation="MALAINSSSSSKAISSSSPSSDLKTKIGSLRLMRTVNSAASLSLGKRSVALNVOSTIKESTIVASEYTERLDVEDEELAKRLENSAPLEIMDKALEFGNDIAIEFSGADVDALIEFAHLTGRIYRFSLDGRNLNPEYLRLEKEKHGRISRYEMPDAAEYQVALVRNGKLFEFYEDGHOECIRIKRVPLRALGLAMITGQRKQSTGTSEIPIVOVDPEFLDGVSILVKMNVPANEGNDVWNEFLTMDPVNTLHAAGVYSICEPCTRAVLPGCHERERGRRMWEDAKAKEGILHKGNLKENTNMATANVGTASADDIFNSENVYNLNRQGIENLMKLENKEAMIVYLIPWCPCOCAMASFEDELADELGSAGKVAKFRADGDOKFAKKELQLGSPFTLVLPKNSSRPITYPSRKVDLSLTSLMLIVR"
/complement(30918..31944)
/gene="AT4g21990"
/number=1
complement(31945..32029)
/number=1
complement(32030..32165)
/gene="AT4g21990"
/number=2
complement(32166..32251)
/number=2
complement(32252..32386)
/gene="AT4g21990"
/number=3
complement(32387..32528)
/number=3
complement(32529..32607)
/gene="AT4g21990"
/number=4
complement(33444..33836)
/gene="AT4g22000"
/number=1
complement(33444..33836)
/gene="AT4g22000"
/number=2
complement(33444..33836)
/gene="AT4g22000"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAI8103.1"
/db_xref="GI:2961345"
/db_xref="SPTREMBL:O65448"
/translation="MAGKAAREAVAKTYVGFOHPRAKIDKTRTELTKVWGCIWEKGANKPLGISARRALKREVLTNGEDWPDERKAMRTKRKGHCODRISAERENTAKMLMKPMQILSRKARKKEWKMEBEKAKEDK"
/complement(37063..37197)
/gene="AT4g22010"
/number=1
37063..40097
/gene="AT4g22010"
join(37063..37197,37310..37419,38362..38632,38736..39184,39290..39658,39806..40097)
/gene="AT4g22010"
/note="strong similarity to pectinesterase, Lycopersicon esculentum, PATX:E312172
contains EST gb:A1994449.1, Z26709, T43750, Z26220,
A1099778"
/codon_start=1
/product="pectinesterase like protein"
/protein_id="CAI8104.1"
/db_xref="GI:2961346"
/db_xref="SPTREMBL:O65449"

Query Match	Best Local Similarity	Matches	Score	DB	Length
Agg1	43.4%	79.7%	53.4	8	103270
Agg2	43.4%	79.7%	53.4	8	103270
Agg3	43.4%	79.7%	53.4	8	103270
Agg4	43.4%	79.7%	53.4	8	103270
Agg5	43.4%	79.7%	53.4	8	103270
Agg6	43.4%	79.7%	53.4	8	103270
Agg7	43.4%	79.7%	53.4	8	103270
Agg8	43.4%	79.7%	53.4	8	103270
Agg9	43.4%	79.7%	53.4	8	103270
Agg10	43.4%	79.7%	53.4	8	103270
Agg11	43.4%	79.7%	53.4	8	103270
Agg12	43.4%	79.7%	53.4	8	103270
Agg13	43.4%	79.7%	53.4	8	103270
Agg14	43.4%	79.7%	53.4	8	103270
Agg15	43.4%	79.7%	53.4	8	103270
Agg16	43.4%	79.7%	53.4	8	103270
Agg17	43.4%	79.7%	53.4	8	103270
Agg18	43.4%	79.7%	53.4	8	103270
Agg19	43.4%	79.7%	53.4	8	103270
Agg20	43.4%	79.7%	53.4	8	103270
Agg21	43.4%	79.7%	53.4	8	103270
Agg22	43.4%	79.7%	53.4	8	103270
Agg23	43.4%	79.7%	53.4	8	103270
Agg24	43.4%	79.7%	53.4	8	103270
Agg25	43.4%	79.7%	53.4	8	103270
Agg26	43.4%	79.7%	53.4	8	103270
Agg27	43.4%	79.7%	53.4	8	103270
Agg28	43.4%	79.7%	53.4	8	103270
Agg29	43.4%	79.7%	53.4	8	103270
Agg30	43.4%	79.7%	53.4	8	103270
Agg31	43.4%	79.7%	53.4	8	103270
Agg32	43.4%	79.7%	53.4	8	103270
Agg33	43.4%	79.7%	53.4	8	103270
Agg34	43.4%	79.7%	53.4	8	103270
Agg35	43.4%	79.7%	53.4	8	103270
Agg36	43.4%	79.7%	53.4	8	103270
Agg37	43.4%	79.7%	53.4	8	103270
Agg38	43.4%	79.7%	53.4	8	103270
Agg39	43.4%	79.7%	53.4	8	103270
Agg40	43.4%	79.7%	53.4	8	103270
Agg41	43.4%	79.7%	53.4	8	103270
Agg42	43.4%	79.7%	53.4	8	103270
Agg43	43.4%	79.7%	53.4	8	103270
Agg44	43.4%	79.7%	53.4	8	103270
Agg45	43.4%	79.7%	53.4	8	103270
Agg46	43.4%	79.7%	53.4	8	103270
Agg47	43.4%	79.7%	53.4	8	103270
Agg48	43.4%	79.7%	53.4	8	103270
Agg49	43.4%	79.7%	53.4	8	103270
Agg50	43.4%	79.7%	53.4	8	103270
Agg51	43.4%	79.7%	53.4	8	103270
Agg52	43.4%	79.7%	53.4	8	103270
Agg53	43.4%	79.7%	53.4	8	103270
Agg54	43.4%	79.7%	53.4	8	103270
Agg55	43.4%	79.7%	53.4	8	103270
Agg56	43.4%	79.7%	53.4	8	103270
Agg57	43.4%	79.7%	53.4	8	103270
Agg58	43.4%	79.7%	53.4	8	103270
Agg59	43.4%	79.7%	53.4	8	103270
Agg60	43.4%	79.7%	53.4	8	103270
Agg61	43.4%	79.7%	53.4	8	103270
Agg62</					

COMMENT	FEATURES
Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/ this fragment has an overlap with ATCHRIV57 at the 5' end and an overlap with ATCHRIV57 at the 3' end.	location/Qualifiers
	1. 196247
	/organism="Arabidopsis thaliana"
	/variety="Columbia"
	/db_xref="taxon:3702"
	/chromosome="4"
gene	5510. 11823
exon	/gene="AT4g21820"
	5510. 6217
	/gene="AT4g21820"
	/number=1
CDS	join(5510. 6217,6310. 6778,7223. 7449,7548. 7613. 7799. 7852,8220. 8336,8544. 8642,8699. 8879,9087. 9224, 9426. 9620,9702. 9982,10872. 11094,11161. 11425, 11469. 11624,11736. 11823)
	/note="similarity to calmodulin-binding protein SH1, Mus musculus, AF062378
	contains EST gb:AI97648.1"
	/codon_start=1
	/product="putative protein"
	/protein_id="CAB79137.1"
	/db_xref="GI:7269027"
	/translation="MDENPPCASPPAPRNPASSLIDISNFKTPRRSYVNSNISKSPYHRTASKOTPKSSSNFRPRSMVHSYASRKYSTSSRRRLKAFELQOSOSRAELTKENRLSLASLTVAWLPLENPENCDCDPENSGVNLGKGRDSEDLGNSKSVGDTMMRSPKRLRLNGWCEKRSSEIDSLGSKYSTLRESLRDVCSDLDLQRMQPHLSLSCKEIFPVMTEVSKNIDEGRIKMKPOCPLVDFGKEKAIKALMSYNQWMLGLYIIFGDSFLDSEVNSDOEMAFKMTVISKQFESHDLARSYYVNNKWEGLRGYYEALGVILKRLILLVLYIDRAKSOSCLSKYIGDIDGSPIMFSEKSKISSHOLICGSEEDMLRVLLKQKVVPSDNKRKRLANCRILQLKLDAGVSLKDEGCMITGEVADGDRELTISLWNMFVHQLPLINGRLLTEETIKVQGVQNNQITMSTPLELWMIQDPGGQEGPQSVASNTDYHDVQNFILSQRLALSLSPFEGDLLEHNAVSNQSVIILLAFSLSKLIVKENDLFTLAGLITNSVSOEQLNHKLICSSQVQEKRSIRISCSSEAVTIEEPRENEEDATKRFQAIKAMMODANONISVGNKNSHTLOGSLRGTDSORAVILLAEIAAVTIOSNIRGLHARRPRMKRAICFLOAAVRTWISVKNIQVVEKFNVEEVTILSERSNLKPARYKFTYDRSFITLRKSVSYIQKAVRRHGSNLHIELKALRIQLAMRSYKEVYISITIOSYVGMVITRRNRRTYKESILIOFOVSYPTISQSHWGYLTRKASKAQVLDLRVMTSAANIDDKRLINKLISLSEL,SMKRVNHLHICETLESATKYSDDKCEELVAAGALEKLLTIRASRSIPDOQVSKHALSTIGHARYPOMADELINTKGSIOITFMELIRSSQVFLSPKNEEAFIADVLKICSSHGKVEAVRKLPALVKRLHVLVEELLRKANIEKRNKVGQTGKEKSERLKEALEYKILITRR"
intron	6218. 6309
	/gene="AT4g21820"
	/number=1
exon	6310. 6778
	/gene="AT4g21820"
	/number=2
intron	6779. 7222
	/gene="AT4g21820"
	/number=2
exon	7223. 7449
	/gene="AT4g21820"
	/number=3
intron	7450. 7547
	/gene="AT4g21820"
	/number=3
exon	7548. 7613
	/gene="AT4g21820"
	/number=4
intron	7614. 7798
	/gene="AT4g21820"
	/number=4
exon	7799. 7852
	/gene="AT4g21820"
	/number=5
intron	7853. 8219
	/codon_start=1

exon	/gene="AT4g21820"
	/number=5
exon	8220. 8336
	/gene="AT4g21820"
	/number=6
intron	8337. 8543
	/gene="AT4g21820"
	/number=6
exon	8544. 8642
	/gene="AT4g21820"
	/number=7
intron	8643. 8698
	/gene="AT4g21820"
	/number=7
exon	8699. 8879
	/gene="AT4g21820"
	/number=8
intron	8880. 9086
	/gene="AT4g21820"
	/number=8
exon	9087. 9224
	/gene="AT4g21820"
	/number=9
intron	9225. 9425
	/gene="AT4g21820"
	/number=9
exon	9426. 9620
	/gene="AT4g21820"
	/number=10
intron	9621. 9701
	/gene="AT4g21820"
	/number=10
exon	9702. 9982
	/gene="AT4g21820"
	/number=11
intron	9983. 10871
	/gene="AT4g21820"
	/number=11
exon	10872. 11094
	/gene="AT4g21820"
	/number=12
intron	11095. 11160
	/gene="AT4g21820"
	/number=12
exon	11161. 11425
	/gene="AT4g21820"
	/number=13
intron	11426. 11468
	/gene="AT4g21820"
	/number=13
exon	11469. 11624
	/gene="AT4g21820"
	/number=14
intron	11625. 11735
	/gene="AT4g21820"
	/number=14
exon	11736. 11823
	/gene="AT4g21820"
	/number=15
gene	12517. 13199
	/gene="AT4g21830"
	complement(12517. 12675)
exon	/gene="AT4g21830"
	/number=1
gene	complement(join(12517. 12675,12834. 12990,13090. 13199))
	/gene="AT4g21830"
	/note="strong similarity to CGI-131 protein, Homo sapiens, AF151889
	contains EST gb:N38012, AA713335, AI996517.1, AA72450, H76478"
	/codon_start=1


```

/product="putative protein"
/protein_id="CAB79138.1"
/db_xref="gi:7269028"
/translation="MTAAVPAATGSGFROKDEEMRAVLSPROFVLPLKGNDRKCKGEF
TKREEGTYSACSCATLTKSTTKFSGCGWPFPAIPQAIKOTPEAGRRMETICA
VCDGLGHVFKGEGYSTPDORHCNVSVSLKSSAGSSQ"
complement(12676..12833)
/number=1
exon
complement(12834..12990)
/gene="AT4g21830"
/number=2
intron
complement(12991..13089)
/number=2
exon
complement(13090..13199)
/gene="AT4g21830"
/number=3
gene
15183..16130
/gene="AT4g21840"
complement(15183..15341)
/gene="AT4g21840"
/number=1
complement(join(15183..15341,15764..15920,16015..16130))
/gene="AT4g21840"
complement(join(15183..15341,15764..15920,16015..16130))

Query Match
Best Local Similarity 79.7%; Pred. No. 6.7e-08; Length 196247;
Matches 63; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 2 aggtattgtaaatgtgagtccttacaacctgatacctaataatgagcgaatgcgagggc 61
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 156910 AGGTACTGCAAAATGTAATGCCCTTACATCCAGACGATCATGCTGCAAGGC 156851
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 156850 TGCAAGACTGTAATGT 156832
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 6
AP003753
LOCUS
DEFINITION
Oryza sativa chromosome 7 clone OJ1339_B08, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION
AP003753
VERSION
AP003753.1 GI:14422471
KEYWORDS
HTG; HTGS_PHASE2.
SOURCE
Oryza sativa (cultivar:Nipponbare) DNA, clone:OJ1339_B08.
ORGANISM
Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartioideae; Oryzaceae; Oryza.
REFERENCE
1 (sites)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
clone:OJ1339_B08
Published Only in Database (2001) In press
2 (bases 1 to 119807)
Sasaki,T., Matsumoto,T. and Yamamoto,K.
Direct Submission
Submitted (13-JUN-2001) Takuji Sasaki, National Institute of
Agrobiological Resources, Rice Genome Research Program; Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@affrc.go.jp, URL:http://rpg.dna.affrc.go.jp/,
Tel:81-298-387441, Fax:81-298-387468)
The nucleotide sequence of this BAC clone was generated by
combining Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.

```

```

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
I. 119807.
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="7"
/clone="OJ1339_B08"

BASE COUNT
ORIGIN
34992 a 23815 c 24131 g 36815 t 54 others

Query Match
Best Local Similarity 42.6%; Score 52.4; DB 2; Length 119807;
Matches 62; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 1 caagttatgtaaatgtgagtccttacaacctgatacctaataatgagcgaatgcgaggg 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 109849 CAGTTCTGCAAGTGTGAGTGCATACATCTGATGACCTTATGATCAATGTAAGA 109908
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 61 ctgcactgactgtaagt 78
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 109909 ATGCTCTGATGTGTGAGT 109926
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

RESULT 7
AF277453
LOCUS
DEFINITION
Arabidopsis thaliana putative PHD finger transcription factor
(SHL1) mRNA, complete cds.
ACCESSION
AF277453
VERSION
AF277453.1 GI:10644644
KEYWORDS
thale cress.
SOURCE
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1062)
Oh,M.-H., Torisky,R.S., Braam,J., Altman,T. and Clouse,S.D.
PHD Finger Dependent Binding of SHL1 to a Specific Promoter Region
of Arabidopsis TCH4 In vitro
Unpublished
2 (bases 1 to 1062)
Clouse,S.D. and Oh,M.-H.
Direct Submission
Submitted (13-JUN-2000) Horticultural science, North Carolina State
University, 228 Kilgore Hall, Raleigh, NC 27695-7609, USA
Location/Qualifiers
I. 1062
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="IV"
/notes="similar to BAC clone F19H22 in GenBank Accession
Number AL035679"
1. 1062
/gene="SHL1"
58..744
/gene="SHL1"
/codon_start=1
/product="putative PHD finger transcription factor"
/protein_id="AA021353.1"
/db_xref="gi:10644645"
/translation="MPKQAPRKQOLSKYLKHNKSIQEGDVAIIRNSSEPKPSYVAR
VEAIFIDARSHAKVAVRWYRPEESIGRQPGAKVEFLSDHPQSDATLEGCK
VHSFSSYTKIDSVGNDPFCRFPEYNSITGAFDDPDRVTPCKCMPYPNPDDIATVQCEC
SEMFHSICITITIEAKKPDNPFCEBSCPOOALHNSNSTSNRDAKYVNRKSLRYTK
SKMKHTRRP"

BASE COUNT
ORIGIN
310 a 207 c 242 g 303 t

```

Query Match	38.2%;	Score 47;	DB 8;	Length 1062;
Best Local Similarity	78.9%;	Pred. No. 9.1e-06;		
Matches	56;	Conservative	0;	Mismatches 15;
			Indels	0;
			Gaps	0;

```

Oy  4  gtattgttaaatgttgagatgccttacaacccctgtagacctaatgycgaatgcgaaggctg 63
    || || || || || || || || || || || || || || || || || || || || ||
Db  477  GTTCTGCAAGTGTGAGATGCCGATTAACCCAGATGACTTGATGTGCAATGCGAGGACTG 530

```

Qy	64	cactgactgct	74
Db	537	TTCTGAGTGGT	547

[illegible]

REFERENCE AUTHORS TITLE JOURNAL REFERENCE	1 (bases 1 to 1058) Speulman, E. and Salamini, F. A barley cDNA clone with homology to the DNA-binding domain of the steroid hormone receptors Plant Sci. 106, 91-98 (1995) 2 (bases 1 to 1058) Speulman, E.

```

FEATURES
source      Location/Qualifiers
1. .1058

```

BASE COUNT	237 a	286 c	283 g	252 t
ORIGIN				

Query Match	36.9%;	Score 45.4;	DB 8;	Length 1058;
Best Local Similarity	77.5%;	Pred. No. 3.3e-05;		
Matches 55; Conservative	0;	Mismatches 16;	Indels 0;	Gaps 0;

OY 4 gttattgttaaatgtagatgccttacaacccctgatgacctaatggcgcaatbcgaagcctg 63
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 539 GTTTTGCAGTGATGATGCCATTACAACTTCAGCATCTTATGATCCAGTGTGGAGTAATG 598

QY	64	cactgactgt	74
Db	599	CTCTGATTGGT	609

RESULT 9

AB011445					
LOCUS	AB011445	367 bp	mRNA	PLN	03-JUL-1999
DEFINITION	Triticum aestivum	WESR4	mRNA for zinc-finger motif,	partial cds	
ACCESSION	AB011445				

ORGANISM
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooideae; Triticeae; Triticum.

REFERENCE	TITLE	JOURNAL	REFERENCE
1 (sites)	Nemoto, Y., Kawakami, N. and Sasakuma, T.		
	Isolation of novel early salt-respondering genes from wheat (<i>Triticum aestivum</i> L.) by differential display		
	Theor. Appl. Genet. 98, 673-678 (1999)		
2 (bases 1 to 367)			

FEATURES	Location/Qualifiers
source	1. .367

BASE COUNT	104 a	76 c	96 g	91 t
ORIGIN				

Query Match	34.0%;	Score 41.8;	DB 8;	Length 367;
Best Local Similarity	80.3%;	Pred. No. 0.00058;		
Matches 49; Conservative	0;	Mismatches 12;	Indels 0;	Gaps 0;

Qy 22 gccctacaaccctgatgacctaattatggcgcaatgcagaggctgcactgaagtgttc 81
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GCCGTACAACCAGATGACCTCATGTGTCGCAATGCCGGGATGCAGAAGACTGGTTCATCC 60

QY	82	a	82
		—	
Db	61	A	61

RESULT 10			
AX059527			
LOCUS	AX059527	44723 bp	DNA
DEFINITION	Sequence 260 from Patent WO0055325.		PAT
			17-JAN-2001

SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots.
Rosidae; eurosid II; Brassicales; Brassicaceae; Arabidopsids.

REFERENCE 1 (bases 1 to 44723)
AUTHORS Preuss, D., Copenhaver, G. and Keith, K.
TITLE Plant chromosome compositions and methods
JOURNAL Patent: WO 0055325-A 260 21-SEP-2000;
The University of Chicago (US)
FEATURES
Source 1. .44723
Location/Qualifiers
BAC COUNT 13679 a 8296 c 8431 g 14317 t
ORIGIN
Query Match 28.0%; Score 34.4; DB 6; Length 44723;
Best Local Similarity 78.8%; Pred. No. 0.31;
Matches 41; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
OY 2 aggtatgttaaatgtgagatccttacaacctgatacctaagtcgcaat 53
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 42002 AGGTACTACAAATGTGAATGCGGTATACACTAGATGACTGATGTCGAAT 42053
RESULT 11
T27D20/c T27D20 80446 bp DNA PLN 03-AUG-1998
LOCUS Arabidopsis thaliana BAC T27D20.
DEFINITION AF076274
ACCESSION AF076274.1 GI:3293583
VERSION
KEYWORDS
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 80446)
REFERENCE 1
AUTHORS The A. thaliana Genome Sequencing Center.
JOURNAL Unpublished (1997)
TITLE 2 (bases 1 to 80446)
AUTHORS Edwards, J., Wollam, C. and Dubbelde, C.
JOURNAL The sequence of A. thaliana T27D20
REFERENCE 3 (bases 1 to 80446)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted by: Department of Genetics, Washington University,
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63108, USA
e-mail: twilson@watson.wustl.edu
MAPPING: Clones were assigned to the YAC map by hybridization by
M. Lodhi, Cold Spring Harbor Laboratories, and fingerprinted
by M. Marra, WashU, to pick the best candidates for sequencing.
NOTICE: This sequence may not be the entire insert of this clone.
It may be shorter because we only sequence overlapping sections
once, or longer because we provide a small overlap between
neighboring submissions.
This sequence was finished as follows unless otherwise noted:
all regions were double stranded or sequenced with an alternate
chemistry; an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one subclone

NEIGHBORING COSMID INFORMATION:

THE 3' CLONE IS T24H24, 200 BP OVERLAP. ACTUAL START OF THIS
COSMID IS AT BASE POSITION 1 OF T27D20, ACTUAL END IS AT 80252 OF
T27D20.

NOTES:
Coding sequences below are predicted from computer analysis, using
the program GeneFinder (P. Green and L. Hillier, ms in preparation).
Location/Qualifiers
1. .80446
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/clone="T27D20"
/chromosome="IV"
/map="unknown"
/complement(12394..13861)
/gene="T27D20.7"
complement(join(12394..12788,13045..13861))
/gene="T27D20.7"
/note="contains similarity to reverse transcriptases
(Pfam: rvt.hmm, score: 12.22)";
/evidence=not_experimental
/codon_start=1
/protein_id="AAC28230.1"
/db_xref="GI:3377848"
/translation="MEGRYCSRDELPRGKTVTSKVFITIKYSDEIEYKRLVAR
GFTQYTGEDLDITFAVAKLHFRVVALSTNTLEMDLQMDVNAFLQDEEYVWK
PPGLGVINAPNPKVFKRAIYGLKOSPRAWYHKLSTTLGRGKTFSEADNTLETPS
OKGIIVILVYVDIIISGNDKVGIDPTKFLKVPDIDKDELKFLGIEVCSKDEL
FLSORKVTIDLLAOGVGLGPKPAKTPLEDYKAKRKGHDNRPFEDATRYRLVGNLY
TMSKQKYVSLSSAERARAKLTTELMLKALKDKDIDTPNPDPHCDQQAIIH
IASNSVFHEKTHIEVDCHKRQVGLPLPHYTESEDLVDFTKGASTKCEYIH
MKLVVDLTRP"
14338..18209
/gene="T27D20.5"
join(14338..14792,14834..14890,14914..15492,15554..16015,
16094..16517,16614..17066,17127..17229,17278..17578,
17636..17983,18077..18209)
/gene="T27D20.5"
/note="contains similarity to reverse transcriptase (Pfam:
rvt.hmm, score: 19.54) and CCHC-type zinc fingers (Pfam:
zf-CCHC.hmm, score: 12.35)";
/codon_start=1
/evidence=not_experimental
/protein_id="AAC2838.1"
/db_xref="GI:3377856"
/translation="MKVEIKTFNGDRDFSPMKIRIPAGLVGLGKNSLDFKLTTP
VKKKEKESEVEDDASDIKOAFPEPDPIKFESEOKAKNTIINTPTVLKVOHCRTA
AEIWLTKLPEMETSLLNRIYTOILKYSKMDVTLSDQVDFELRIADCLDQNLTP
SSYIOLKHLNDVSSASLERKLESESNKNVSMALYTDGRPPQVRNDKQOG
KNGRSNSTRVYCMFCCKEHAHYKRCDFARKKLENNRATMLLDHDIYERSGCT
IKNTHGSLIRMLKNRYRVYENLRNLISGTLHSLGKHGEGKLFYKNGKTAGCT
YLMNGLIYLDGHTVAETPCNANSAKSNSTLIEKFEKFDCEEHVPMKSKLSTNV
GKHVPDILIGYHADLMGSPNTPPSISGQVPLSTIDDSRKWMLFKSKDTPPKF
CEKELTEVNHVGKVKTLRTDNGLEFCNNRPNFYDKCTGIERRTCTYTPDQNGVTR
MNTTIEKVRSPASADVHNVPQMLNRPQYKHLRFSSIAVHDEQKLRKRYLKG
VFLEGPQGTGRKXVMDIDEKVCISRNVPENEDSVFKDQSGSKDEDEYTOQSQIEI
EPPKSELTQNOVGATOVOLNDESEDEYNGVEGPAFLVAEYSESEPPFHPVK
EPKDEKMHGMIEMDLSLKNATMDIVKPKNOKYISCHMLVKKLGIIGVLTLPKY
AHLVARGFSHREIDIOEYFSPVYKTSRIILSLMKEDMELEQMDVETAPLHGELD
QTLMEQPEGEAFPRQWNRKFNAPMDQKFSRSQVSCYVKEVSNASKNTEIKRLK
VLSREFEMGDMGAASRLGIDILIRNSSEGTLCISQVSYLEVYQKRMGAKRVNPIPI
GAHFKLSSVHNDERVSERVPYSSVSGSLMEPSKQGEVHWMAVKMLRYLKSLIGLN
LMTKFDKVOGCDSDHAADLIDKMSLSGYFTYVGGNIVSKSCLQVLAISTTEA
EYIALRKAVKEMWIRNLDDMLGTEFAVEDVITIEGVEVHKIHTTRNPDMLTFCGI
PYSKPEALAEFLKLR"
22634..25080
/gene="T27D20.4"
join(22634..22705,22988..23235,24132..24348,24458..24691,
24826..25080)
/gene="T27D20.4"

/note="contains similarity to the basic region plus
leucine zipper transcription factors (Pfam: bzip_hmm,
score: 12.77)"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC28227.1"
/db_xref="GI:3377845"
/translation="MAGGKGGSSGSRGSRGSTRPAKLADPPETQIPLKQNHHEE
DEEDGEDREVEGEDEEVDEKQNPVADYQELDLRLALPGROSLVLSQEPPE
VETLMKGIIVSAKSGSEOPWIRDTLKMOMVHNHTEDPAQKSETSNCRSDRGL
GVHKLHAGOKSPVOYHOMERLKRHVSQGEVMOTHTRYDSFVQOKSGHGEAVYKI
LEERSQEMDEGEPQYDSNQSOSTHRTILIDKNNITFLCTHIDKGNPFGLSULET
LNKRKRKESYVSSSTVVAQLQELQQLKISEQEQNAKIDEEHROSQSIASLEKLV
FMKERS"
complement(27921..28630)
/gene="T27D20.8"
complement(join(27921..28322,28562..28630))
/note="contains similarity to DNA
(cytosine-5-) methyltransferases"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC28226.1"
/db_xref="GI:3377844"
/translation="MAKTRPGVAFSSKIKLCKDKILGRELILGNVYEVFFYVG
CVLMPSDAGKAPYARVEKLEADARNNNKVMYICEESHGGRQJLHAKLEFLS
DHFVQSAHTLEGKCIVHTFKNYTRLNENGVBDYCIPEYKATGAFPPDRAV"
35430..36929
/gene="T27D20.3"
join(35430..35907,35990..36096,36168..36929)
/gene="T27D20.3"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC28229.1"
/db_xref="GI:3377847"
/translation="MASSSYCAEKKMERERAREKERSLOSDKVVMAGTLEDKRLAD
KAKOPNEDDSGTROKSSALSGPKISGSEKRTNKTNPDEPKRRDARKKRSER
DARSPRSRRGNENKKEGTQONQKKNKGSODLVYSSRSERISQRTSTPLPA
PPMPADTLRLVHPRAIMPLAEIKATNREKLGEAVENYAFLSHEDQLEKKEI
ESYKQNEENARMWRANSVLSRMRAEAQVOTLEISNTDLSAKLESQNAVPTIEN
ENRSFAELACEERMKLEEGQSMHAAROEKRRVROLTFESKYYCTFVQSDVEV
EAANIRAEVAVNRELLIEIKELIPSTAELESYRADEIKFAQAAAEKTRPPNVE
PTSLADTPPEPVINDSNASDEVLILDKICTNKTQNAEAGIAMFPYDVAKNEAAE"
40097..41244
/gene="T27D20.19"
join(40097..40331,40377..40461,40515..40592,40638..41244)
/gene="T27D20.19"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC28236.1"
/db_xref="GI:3377854"
/translation="MATLTIELRGRLRRKRLGLEQPYPRVELGODQARRNAREEAEFP
DLDEDDPPDPLRQHOQREDPLRNARAGVEPRKSNPEVMDMERREYEFEECYGM
ENLGRVADRRRRKRGPIINTGDMKYLRLRLSOGTRVDEPFEPEFKLMALTEE
SDEALMAQFIDGLOERIORKVEAOYNGLHLLHAYOEQIRRKASISNRRNTP
WNASNNRMDKSKVYESDHRKKNKSNPKTSRPLGFPSTNQSRSRITTFKQGR
GHMARECLNQRMVILTPSGDYEQDEDSDEHEDIEYPDVGLLVIRRTLVSFVNLE"
41442..42505
/gene="T27D20.2"
join(41442..41817,41934..42505)
/gene="T27D20.2"
/note="contains similarity to reverse transcriptase (Pfam:
rvt_hmm, score: 96.80) and CCHC-type zinc fingers (Pfam:
zf-CCHC_hmm, score: 14.43)"
/codon_start=1
/evidence=not_experimental
/protein_id="AAC28237.1"
/db_xref="GI:3377855"
/translation="MRVGHLLGRPMQFDRATCNHRTNHYSFYNDKRYNLAPLSL
EVHDLQIHMKNEHEVRSSLYLSKQNKYSAKTYVLMFKEKLSIGIDPELSAE
VOVVLHMYKDLFPEEIPGGLPIHKGYSRESISPAVPLVLPKKGQWRMCLDRAI
NNITIKYRHPILRPLYMDLDELGAIIFFSKVDLRSGYHQVRMKEGDEMKTAKTKOGLY

gene
ECLVMPFGLTNAPSTFMRLNQLNQLRSFGKRVVYFEDILYKNSYSDHIOPLELILK
TLRKEGLYANKKCTGCSDEKFEFPICVDFVNNILLIIN"
complement(46940..49459)
/gene="T27D20.9"
complement(join(46940..48645,48865..49459))
CDS
Query Match 28.0%; Score 34.4; DB 8; Length 80446;
Best Local Similarity 78.8%; Pred. No. 0.32;
Matches 41; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 2 aggtattgtaaatgtagtagtgccttaaacccctgtagcctaagtggcaat 53
||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1 ||||| 1
Db 27838 AGGTACTACAAATGTAATAACCGATATCTAGATGATGATGATGAT 27787
RESULT 12
ATCHRIV11
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 197052 to 198022)
Zhong,J., Ma,P., Parnell,L.D., Chen,C.N., Chen,E.Y., Mewes,H.W.,
Lemcke,K. and Mayer,K.F.X.
Unpublished
2 (bases 1 to 28786; 28587 to 117434; 117235 to 197680)
Mayer,B., Stoneking,T., Stumpf,J., Mewes,H.W., Lemcke,K. and
Mayer,K.F.X.
JOURNAL
REFERENCE
AUTHORS
JOURNAL
AUTHORS
EU Arabidopsis sequencing, project.
3 (bases 1 to 198022)
Unpublished
Direct Submission
Submitted (10-MAR-2000) MIPS, at the Max-Planck-Institut fuer
Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
lemcke@mps.biochem.mpg.de, mayer@mps.biochem.mpg.de, Project
Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge
Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,
E-mail: michael.bevan@bsrc.ac.uk
Information on performance of analysis and a more detailed
annotation of this entry and other sequences of chromosomes 3, 4
and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/
this fragment has an overlap with ATCHRIV10 at the 5' end and an
overlap with ATCHRIV12 at the 3' end.
FEATURES
source
Location/Qualifiers
1..198022
/organism="Arabidopsis thaliana"
/variety="Columbia"
/db_xref="taxon:3702"
/chromosome="4"
8288..8638
/gene="AT4G03970"
/number=1
8288..12735
/gene="AT4G03970"
join(8288..8638,8738..8949,8993..9236,9341..9454,
9527..9676,10448..10561,10690..10770,10833..10905,
10956..11074,11199..11279,11426..11509,11626..11679,
11793..11986,12069..12153,12238..12735)
/note="AT4G03970"
/note="contains similarity to human CD44R1 gene product
(GB:X56794)
similarity to"
/codon_start=1
/product="hypothetical protein"
/protein_id="CAB77865.1"
/db_xref="GI:7267153"
/translation="MLADLDKFLSPWGRESFLKTAGMBRDKRLGKSTGKROLTTD

PIKSLVSOLOQKTFERLKGFPFPALQLIAPRNIPQLDLVLDDSTSKTILMKRLTHPKQ
VISLPQIHALSENLKLPVDPPLIYGDNLNADQNAEMDEYKDKVAYMYDLIWSHOFQ
KIDWPGGADATLPIISVKKKOSRGCGPRAYSRSPKPELTINLSDPEKSRILISELOTVE
ELSNRPYMKLEKAKKTVRKRRTKSSIFACSSISKRKIMEYPIIOSOSESTLPEH
VFGQPEVAAQVPPDDHLDTKODSSDSIPLINTEITDDPMVFVTPLDSESNKDDANEG
NVPYDITDVKDQANENEDVDSQMODASERVPETHSGLDLPKEHSELETNANETDVG
KQODSLDRETASHSHIDLPIDQSSSEQAKIMSELIMVFKDYNLMSLETINFAANNIY
LCLSHIIFEYVNIQDITMEAAASHSLGPAKPAKKVTLESAVSQLSQTPEPEYFD
DDPMKAMVYIYSHPHDNTSKPNKOLESSPTPEEDOTDSDIGAHDLDENOEGYVDV
SDSPAREREKPTLSAEVYELVAELLSKSGTGYSELLPSMSKSEFALPFRNTLSAAPT
EHITSSGYLISNKKFLSLAKPTINWSTLMEVYVLSLQKLATTLTNQRAAYQPFPA
NHLQGLKSFKAAMKSRVKSWEPMKQFTVGPTEWFADITLYLPMIWDSKHWGLA
INLGVMSVEILDPNTDLYEDEVRRIEPPVYTIMPYLIORYCPECSQNHGLOPFYWK
RLDGLTKNLRSGCGPVAM"
8639..8737
/gene="AT4g03970"
/number=1
/gene="AT4g03970"
/number=2
8950..8992
/gene="AT4g03970"
/number=2
8993..9236
/gene="AT4g03970"
/number=3
9237..9340
/gene="AT4g03970"
/number=3
9341..9454
/gene="AT4g03970"
/number=4
9455..9526
/gene="AT4g03970"
/number=4
9527..9676
/gene="AT4g03970"
/number=5
9677..10447
/gene="AT4g03970"
/number=5
10448..10561
/gene="AT4g03970"
/number=6
10562..10689
/gene="AT4g03970"
/number=6
10690..10770
/gene="AT4g03970"
/number=7
10771..10832
/gene="AT4g03970"
/number=7
10833..10905
/gene="AT4g03970"
/number=8
10906..10955
/gene="AT4g03970"
/number=8
10956..11074
/gene="AT4g03970"
/number=9
11075..11198
/gene="AT4g03970"
/number=9
11199..11279
/gene="AT4g03970"
/number=10
11280..11425
/gene="AT4g03970"
/number=10
11426..11509
/gene="AT4g03970"

/number=11
11510..11625
/gene="AT4g03970"
/number=11
11626..11679
/gene="AT4g03970"
/number=12
11680..11792
/gene="AT4g03970"
/number=12
11793..11986
/gene="AT4g03970"
/number=13
11987..12068
/gene="AT4g03970"
/number=13
12069..12153
/gene="AT4g03970"
/number=14
12154..12237
/gene="AT4g03970"
/number=14
12238..12735
/gene="AT4g03970"
/number=15
20580..22438
/note="pseudogene"
26352..28320
/gene="AT4g03980"
26352..26944
/gene="AT4g03980"
/number=1
join(26352..26944,27053..27955,28074..28320)
/gene="AT4g03980"
/note="similarity to similar to maize transposon 'mudr
mudra protein (GB:AL021710)"
/codon_start=1
/product="putative transposon protein"
/protein_id="CA877866.1"
/db_xref="GI:7267154"
/translation="MFKVSFLFWQPLRTSRIIOACVSEMSSTIFDQYSSSAFSTGLSD
ADSLFEGKKEKDKDEMYFTLRMFVAVHSEPFHTVKSDLTRYVLACIDENCQMKRLRATK
AGSESESYVIRKTVSHHSCDSLRLNYSHRQAFATLGLRLSNHLEGKLPDGPQLIET
FRDHGVGINYSKAMRVOEHVVELARGLPDDSPREVLPBGYKILRKRYISIDGAHLTSEF
KGTILGASADQGNENLXPASAFATVDENASDMFLPKCLNITIPDNDLVFVSDRAAS
IASGLSENPLAHGICLTFHLQKNLTFHRGSSLIPVYNVAAASRYTKTERDESLFKKIT
NSDKRLAQYLMEVDVWRKWSRAYSPSNRYNIMTNSLAESVNALIKQNRREPIVCLFESI
RSIMTRFMNERRESSQHPASVATINVGKMKASDYDSTFWLVEVCQYQOEFEVKGDTK
THLVINDKRTCTCCMEDIDKFPYAHGIAAKHINLKNMFEVDEHSTYRMROYSESI
HPNGDMEYWEIPDLYSEVICLPSTRVPSGRKKRKRIPSVMEHGRSQPKKLHKSCRC
GSGGHNKSTCVAAI"
26945..27052
/gene="AT4g03980"
/number=1
27053..27955
/gene="AT4g03980"
/number=2
27956..28073
/gene="AT4g03980"
/number=2
28074..28320
/gene="AT4g03980"
/number=3
31256..32611
/gene="AT4g03990"
complement(join(31256..31513,31718..31843,31964..32611))
/gene="AT4g03990"
complement(join(31256..31513,31718..31843,31964..32611))
/gene="AT4g03990"
/note="similarity to"
/codon_start=1

Query Match

28.0%; Score 34.4; DB 8; Length 198022;

Best Local Similarity 78.8%; Pred. No. 0.34; Matches 41; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 2 aggtatgttaatgtgatgaccttaaacacctgatgacctaatgagcgaat 53
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 169843 AGGTACTACAAGTGAATGCCGTATTACTCAGATCAACTGATGCGCAAT 169894

RESULT 13
AL161773/c AL161773 154195 bp DNA PRI 17-JUL-2001

LOCUS Human DNA sequence from clone RP11-212E4 on chromosome 13, complete
DEFINITION sequence.

ACCESSION AL161773
VERSION AL161773.21 GI:14970659
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS Peck, A.
TITLE 1 (bases 1 to 154195)
JOURNAL

COMMENT Direct Submission
Submitted (17-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 19, 2001 this sequence version replaced gi:13751315.

During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/projects/C-elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-212E4 is from the library RPCR-11.1 constructed by the group
of Pletier de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone
RP11-212E4. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true right end of clone RP11-212E4 is at 154195 in this
sequence. The true left end of clone RP11-23116 is at 133478 in
this sequence. The true right end of clone RP11-472K17 is at 2000
in this sequence.

FEATURES
Source Location/Qualifiers

1..154195
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-212E4"
/clone_lib="RPCR-11.1"
346..646
repeat_region
/note="AluX repeat: matches 1..302 of consensus"
1158..1196
repeat_region
/note="L2 repeat: matches 2693..2731 of consensus"
1415..1720
repeat_region
/note="AluY repeat: matches 2..298 of consensus"

repeat_region 3397..3622
/note="MER20 repeat: matches 1..218 of consensus"
repeat_region 6907..7214
/note="AluX repeat: matches 1..308 of consensus"
repeat_region 9368..9663
/note="AluX repeat: matches 1..295 of consensus"
9761..9869
/note="MIR repeat: matches 120..244 of consensus"
repeat_region 13836..13879
/note="L2 repeat: matches 2632..2692 of consensus"
repeat_region 13883..13918
/note="9 copies 4 mer aggg 86% conserved"
14560..14837
/note="AluX repeat: matches 1..277 of consensus"
15134..15378
/note="LIME3 repeat: matches 5895..6144 of consensus"
15675..15791
/note="LIME3A repeat: matches 6034..6158 of consensus"
17512..17569
/note="L2 repeat: matches 2632..2692 of consensus"
18079..18368
/note="AluX repeat: matches 1..290 of consensus"
18409..19128
/note="12 copies 60 mer 59% conserved"
18514..18678
/note="5 copies 33 mer 83% conserved"
18676..19822
/note="37 copies 31 mer 56% conserved"
18721..18756
/note="9 copies 4 mer atat 83% conserved"
18831..19282
/note="14 copies 33 mer 89% conserved"
19265..19534
/note="9 copies 30 mer 60% conserved"
19267..19330
/note="32 copies 2 mer ta 78% conserved"
19510..19689
/note="3 copies 60 mer 73% conserved"
19553..19849
/note="9 copies 33 mer 85% conserved"
19821..19872
/note="26 copies 2 mer ta 75% conserved"
19823..19882
/note="3 copies 20 mer 85% conserved"
19874..19925
/note="26 copies 2 mer ta 76% conserved"
20002..20114
/note="19 copies 22 mer 89% conserved"
20176..20241
/note="3 copies 22 mer 89% conserved"
20925..21135
/note="L2 repeat: matches 2490..2708 of consensus"
21566..21983
/note="19 copies 22 mer 61% conserved"
21571..21990
/note="14 copies 30 mer 63% conserved"
21573..21980
/note="17 copies 24 mer 64% conserved"
21589..21932
/note="8 copies 43 mer 73% conserved"
21593..21840
/note="8 copies 31 mer 71% conserved"
21600..21995
/note="11 copies 36 mer 62% conserved"
21609..21992
/note="12 copies 32 mer 64% conserved"
21668..21907
/note="12 copies 20 mer 77% conserved"
21696..21995
/note="5 copies 60 mer 72% conserved"
21723..21953
/note="7 copies 33 mer 69% conserved"
21755
misc_feature

Sequence updated (26-May-2000).

* NOTE: This is a 'working draft' sequence. It currently consists of 16 contigs. The true order of the plasmid is not known and their order in this sequence represents an arbitrary. Gaps between the contigs are represented by runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number be preserved.

16517	162626	contig of 1070 bp in length
1	43657	contig of 43657 bp in length
43658	43757	gap of 100 bp
43758	60093	contig of 16336 bp in length
60094	60193	gap of 100 bp
60194	77413	contig of 17220 bp in length
77414	77513	contig of 100 bp
77514	94208	contig of 16655 bp in length
94209	94308	gap of 100 bp
94309	107738	contig of 13430 bp in length
107739	107838	gap of 100 bp
107839	117065	contig of 9227 bp in length
117066	117165	gap of 100 bp
117166	124047	contig of 6882 bp in length
124048	124147	gap of 100 bp
124148	131440	contig of 7293 bp in length
131441	131540	gap of 100 bp
131541	138828	contig of 7288 bp in length
138829	138928	gap of 100 bp
138929	144091	contig of 5163 bp in length
144092	144191	gap of 100 bp
144192	150955	contig of 6764 bp in length
150956	151055	gap of 100 bp
151056	156185	contig of 5130 bp in length
156186	156285	gap of 100 bp
156286	158976	contig of 2691 bp in length
158977	159076	gap of 100 bp
159077	162099	contig of 3023 bp in length
162100	162199	gap of 100 bp
162200	165096	contig of 2897 bp in length
165097	165196	gap of 100 bp
165197	166266	contig of 1070 bp in length.

[illegible]

```
misc_feature /note="assembly_fragment clone_end:sp6 vector_side:left
159077..162099 /note="assembly_fragment"
misc_feature 163200..165096 /note="assembly_fragment"
misc_feature 165197..166365 /note="assembly_fragment"
BASE COUNT 47162 a 37047 c 36491 g 44059 t 1507 others
ORIGIN
```

Query Match	27.5%;	Score 33.8;	DB 2;	Length 166266;
Best Local Similarity	56.9%;	Pred. No. 0.55;		
Matches	62;	Conservative	0;	Mismatches 47; Indels 0; Gaps 0.
Qy	8	tgtaaatgtagatgcttcacacccgatgacccaatgagcgcaatgcgagggctcact	67	
Db 124016	TATACATATTAATAATATCATATATTAATATTAATTAATTCGATTTGCCCTCAATGTTATACAGACT	123957		
Qy	68	gactcgtgaagcttcacatcgaagcgatcaatcattcaattcaatca	116	
Db 123956	GTGCTTTAAATTAACATCATATGTGTGCTTTATTTTAAACCTATTTCCA	123908		

RESULT	15
AC025409	
LOCUS	AC025409 185404 bp DNA HTG 08-NOV--2000
DEFINITION	Homo sapiens chromosome 1 clone RP11-304111 map 1, WORKING DRAFT
SEQUENCE	SEQUENCE, 55 unordered pieces.
ACCESSION	AC025409
VERSION	AC025409.3 GI:10047870
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo. 1 (bases 1 to 185404)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens chromosome 1, clone RP11-30411
Unpublished
2 (bases 1 to 185404)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, J., Barua, N., Bastien, V., Beda, F.,
Bogdanovskiy, L., Boukhaltter, B., Brown, A., Burkett, G.,
Campopiano, A., Castle, A., Choeplou, Y., Colangelo, M., Collins, S.,
Collymore, A., Cooke, P., DeArrellano, C., Dewar, K., Diaz, J. S.,
Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,
Galagan, J., Gadyana, S., Ginde, S., Goyette, M., Graham, L.,
Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
Klein, J., Larocque, K., Lamazars, R., Landers, T., Lehoczy, J.,
Levine, R., Lien, C., Liu, G., Locke, K., Macdonald, P., Marguis, N.,
McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheteters, R.,
Melchidim, J., Meneus, L., Mihova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Olyar, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Ryan, A., Santos, R., Schauer, S., Severy, S., Sudranthan, A., Talamas, J.,
Stange, Rihmann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tessilaye, S., Theodore, R., Tirrell, A., Travers, M., Triaillo, J.,
Vassiliadis, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

TITLE
JOURNAL

COMMENT

Submitted (08-MAR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 10, 2000 this sequence version replaced g1:731563.

Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>


```
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 181126
Center clone name: 304-111
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 158038 bases at least Q40
Consensus quality: 173087 bases at least Q30
Consensus quality: 177952 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 180000; sum-of-contigs
Quality coverage: 3.8 in Q20 bases; agarose-fp
Quality coverage: 3.3 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 55 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 9433: contig of 9433 bp in length
* 9434 9533: gap of 100 bp
* 9534 10744: contig of 1211 bp in length
* 10745 10844: gap of 100 bp in length
* 10845 12149: contig of 1305 bp in length
* 12150 12249: gap of 100 bp
* 12250 13273: contig of 1024 bp in length
* 13274 13373: gap of 100 bp
* 13374 14552: contig of 1179 bp in length
* 14553 14652: gap of 100 bp
* 14653 15679: contig of 1027 bp in length
* 15680 15779: gap of 100 bp
* 15780 16942: contig of 1163 bp in length
* 16943 17042: gap of 100 bp
* 17043 18595: contig of 1553 bp in length
* 18596 18695: gap of 100 bp
* 18696 19867: contig of 1172 bp in length
* 19868 19967: gap of 100 bp
* 19968 20972: contig of 1005 bp in length
* 20973 21072: gap of 100 bp
* 21073 22260: contig of 1188 bp in length
* 22261 22360: gap of 100 bp
* 22361 23548: contig of 1188 bp in length
* 23549 23648: gap of 100 bp
* 23649 24920: contig of 1272 bp in length
* 24921 25020: gap of 100 bp
* 25021 26233: contig of 1213 bp in length
* 26234 26333: gap of 100 bp
* 26334 27853: contig of 1520 bp in length
* 27854 27953: gap of 100 bp
* 27954 29293: contig of 1340 bp in length
* 29294 29393: gap of 100 bp
* 29394 53904: contig of 24511 bp in length
* 53905 54004: gap of 100 bp
* 54005 55149: contig of 1145 bp in length
* 55150 55249: gap of 100 bp
* 55250 56417: contig of 1168 bp in length
* 56418 56517: gap of 100 bp
* 56518 57905: contig of 1388 bp in length
* 57906 58005: gap of 100 bp
* 58006 59297: contig of 1292 bp in length
* 59298 59397: gap of 100 bp
* 59398 60783: contig of 1386 bp in length
* 60784 60883: gap of 100 bp
* 60884 62491: contig of 1608 bp in length
* 62492 62591: gap of 100 bp
* 62592 64008: contig of 1417 bp in length
* 64009 64108: gap of 100 bp
* 64109 65352: contig of 1244 bp in length
```

```
* 65353 65452: gap of 100 bp
* 65453 66782: contig of 1330 bp in length
* 66783 66882: gap of 100 bp
* 66883 68395: contig of 1513 bp in length
* 68396 68495: gap of 100 bp
* 68496 70756: contig of 2261 bp in length
* 70757 70856: gap of 100 bp
* 70857 72429: contig of 1573 bp in length
* 72430 72529: gap of 100 bp
* 72530 73566: contig of 1037 bp in length
* 73567 73666: gap of 100 bp
* 73667 74917: contig of 1251 bp in length
* 74918 75017: gap of 100 bp
* 75018 77012: contig of 1995 bp in length
* 77013 77112: gap of 100 bp
* 77113 79544: contig of 2432 bp in length
* 79545 79644: gap of 100 bp
* 79645 81091: contig of 1447 bp in length
* 81092 81191: gap of 100 bp
* 81192 83043: contig of 1852 bp in length
* 83044 83143: gap of 100 bp
* 83144 85037: contig of 1894 bp in length
* 85038 85137: gap of 100 bp
* 85138 87330: contig of 2193 bp in length
* 87331 87430: gap of 100 bp
* 87431 89273: contig of 1843 bp in length
* 89274 89373: gap of 100 bp
* 89374 90847: contig of 1474 bp in length
* 90848 90947: gap of 100 bp
* 90948 93191: contig of 2244 bp in length
* 93192 93291: gap of 100 bp
* 93292 95994: contig of 2703 bp in length
* 95995 96094: gap of 100 bp
* 96095 99821: contig of 3727 bp in length
* 99822 99921: gap of 100 bp
* 99922 103654: contig of 3733 bp in length
* 103655 103754: gap of 100 bp
* 103755 107159: contig of 3405 bp in length
* 107160 107259: gap of 100 bp
* 107260 110785: contig of 3526 bp in length
* 110786 110885: gap of 100 bp
* 110886 115380: contig of 4495 bp in length
* 115381 115480: gap of 100 bp
* 115481 119514: contig of 4034 bp in length
* 119515 119614: gap of 100 bp
* 119615 123779: contig of 4165 bp in length
* 123780 123879: gap of 100 bp
* 123880 133095: contig of 9216 bp in length
* 133096 133195: gap of 100 bp
* 133196 139671: contig of 6476 bp in length
* 139672 139771: gap of 100 bp
* 139772 147743: contig of 7972 bp in length
* 147744 147843: gap of 100 bp
* 147844 156603: contig of 8760 bp in length
* 156604 156703: gap of 100 bp
* 156704 169229: contig of 12526 bp in length
* 169230 169329: gap of 100 bp
* 169330 185113: contig of 15764 bp in length
* 185114 185213: gap of 100 bp
* 185214 185404: contig of 191 bp in length.
*
* location/Qualifiers
* 1. 185404
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="1"
* /map="1"
* /clone_1lb="RP11-304111"
* /clone_1lb="RP11 Human Male BAC"
* 1. 9433
* /note="assembly_fragment"
* clone_end:SP6
* vector_side:left"
* 9534 .10744
```

FEATURES

source

misc-feature

misc-feature

gene
3255..13022
CDS

/gene="TC0437"
/note="similar to GP:6013469; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="adherence factor"
/protein_id="AA039291.1"
/db_xref="GI:7190480"

gene
13043..23050
CDS

TVIGSANSWNSGRLNTNKEDLLIISPLTEGOKODICTWMHFLCENSTVLAEEKTPKI
LD"
3255..13022
/gene="TC0437"
3255..13022
/gene="TC0437"
/note="similar to GP:6013469; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="adherence factor"
/protein_id="AA039291.1"
/db_xref="GI:7190480"
/translation="MPPSAISKISNSTTPVOYSSAPNLTMLLEGKISVEKSRVSE
EENQHTTEDVAVRICAMLOOQFYNNMTGRNESTVSPAHSGMNTSLIYNQAQVAH
IFSLNTYPAKIIHHNGTKOMNESKOPVSSSEKSNKRNVNSESIAVONENVEVTS
SLAKTSIRRRKPREVEERKQOLVDPADKIIYDLRADPIASRLKTEOIOOKISL
NLIKATORYNIDSEBNSRSGCEDLYKQSFILNNILENAMESSEBKNNVSIKT
EPNSHRVIDSHFHGTWITAGAPPECTDVYIKTFLLDTIEFDLEWVDRQAFANASS
ILKRIAFDSSSLKELKRSITSQETQDVKKEYDQAKYQTSGRREERLELEKRLKLEDR
YKHLQSEIRTNEDALFLKNVLSODGFENYCLKIGTIDQTRAYLQDELHLPEE
IOQYKDLKANEOKIKDIYGVNTDGLKERYIKNIRLDSMODRTNRYNEMMLR
WNPASAPQIRMYMLKELGGIYTDLDMPQYSPVLOMINEIGDGFPELPITRAVS
DGVRLANGESGITTDNIADODISKILRSDPOLEKLTIDENOKSSGSESLFO
RMADSVADPMPILQRIHKSTGWNVRLNGIMAHKKSAMADANIKGQRAVRLKS
LRETVLSEEFRTLDLKLHNKALIGHLVEFLGKSLFSEFRODTIIAGLSTLGI
TGPDLYDEMERTYRSIGPIGKLDLEKRLGKAPLGSYKLVKEGKOTFDMLHVTY
GANVTPADESTWGIQORLAELFSDSEKFRPTPKGISRTKIDKFTFLMEOS
KRALSPELDRENFELIEAKNFDLIKFTVDRLAYLRNOVCDAASAVAFLQOLA
ELRGTSPIPIANOVNPPDLNRNTIASSEKTKIKLYLOSHQSEYVIMSAISDRVMP
RDMLATARMNLVKESTSEKTEPTLTTEIKILNSMAELQSKDNIGILSEVDKLD
VTSIEMSKLKRVSALIEDSTVAGTFYLRLEETISEWRIPEKDLKRVIREMGMD
ANQHPDERGRWYSQLYDESFKRKVTAPVKRIQELAKVYLMEOVHLEIDYLTKN
PLFRLHEEGYAFSDLEITRYMLAEYVIGSIFSGNLLPSPARLVNI IKYVGDY
HDMQDLVKIYDMLASGAGADLTNERFAIPESLRNLEGLTDLPLPPVAVSGW
GMSGFENGVEGSDMLMISIAQFENGASYSMQHYALYEIHRHILGSLTDLKKE
LESKAGCFVHEERDSSLKASSEKQSLTEIHKISNOVHLAEVASHMNTALPY
GKITERBEDFRGLCTTTEYVAMRSYTRIGIYSKDLFSLPEVPTIONVEQAKY
LLSWPEFYNRANAMASDLATHYGAELILAHPOSFLYEVEGRCMGSLJLMEDEGGY
RTLOGNDLTVALYQOKERDHLPLSKQDLSLRDLSLIMLOYQGNVLEENRPNH
AKMVLQITSTFEKSLTSLITTPTHSLTIFMGSFPRVDPNGHVDLLEAFNL
FIEDMVOYSPHIOKRGFSESLTSVREOLQVYFLDEPQISLSTFDGLSHOAT
LEKMBRGVPIERIOOTWGTLYETIGSSIHHRRIDEPTSEIDPMKINGDVLPEIT
KHVIDSEYSLILVLETRGLEBGRKHYSRSLIYTPPEASLFOPLNKRKAKTEML
VSLSEVYDRLKSQOLDEDEVEVSRSVSDSDTEVEKKGSLKLKLSIRGGJFDS
FKQPSAKNELGATGVMDLEGMSIVSYOVNRLIEBKGKGPLANVIMVOKETAE
MTGAVIQLGSKKETQBGVDFRLETAIASOLKASSRVGSGVSKTAAVLELP
VLETAIGAMNLYDSVYSLKEAOTHSDRMAARVDAFNSITLIGITIASVAPLSIMAVG
PLAIGMASAIARNVAKREBYDMVLYKKFTGSGSHILFADPDRLHLDSSGYVL
GNFLDLRQNPPIIKGDRSYNSNRITGKAGNTMOIKRIGYGRISPTSLARGHA
NSFWPELPPIPKGDYRVILGYTTYEAVTEVYLSNKKVYREAVMSSSXYRPPPL
TAKSKLSIIIGDSDPLVYIPVRLIEKDEKENIDHNASYKDKINIKGKGGLTVAIG
AGFYNLGTLPATNTISPRAPVAPGFENSLVODVPLRPNCTEKKFLIRDTG
NTILGSEGNDTLGGNDTKFYSSPGGTYISGKGVHEYHLPOLPHIILISKSSA
HOLYTEINAQOVKPYLGSLSLTSSEMLISKOEIRHSSDETKEDLGWKKEREVRLGD
ITATPIDMAGTRVLVWGVTCTDAAKQKHPIETSTPEDLDMLTOKSKWEVRLGDL
FVLENGTAKFTNKRHIYSPKPYSELAQTOSRKAVQNGTSGVFSYIYRSQK
TOQTELTLYDDVDAQOTIDVSSILPTLLAKSGDNQIDFEFSSHHMLAFISRSQ
IIPERTRVKINNETILRLGMLNAKLMASSNPYNNLNYNRNIPERVADVSLNNTV
LMPRTNRNGEHVAIEMQEGVDVYKFGLOSCHVLEEDIMVAKYDPAKITHFAV
TYPAHKTKYLEFTDEESLANPEDVNLPSKIGSNTLOASQKPLTDFANQMLSTDEIS
VSDTSLENFVRYLMEAEETIELYRQLLYAONLNLNHRDVLKLEFVYRQSGKMAVL
ITFRNFIEMSEMGISAPTLEREAKPIILASDNGPIYREHLDLEFEDKMLNLAIV
QEFMLOIYVIMQEGKALGSLTTPRCMDMLAVLYVAVSSMEGRASMLAKAIEY
KLPELTTLTASYLYXVSGDLVYPCIMNRPRONQAVYFRFGYKQKOPKAPSIATSS
HSVGESRTAKVLOPFOQELFHLKSANRKAVALQOLDSKLSLSMLLYGGDQVIRY
PLVQOQPLSDDLMADLKERVYMANNEGFLISOMHSHLIMESOMNLRDVPKIN
PNVINNAITYYRVVAVPEWIRNRLTKGSLIRIPARSHIPALTTOODVKTLYAPESL
YFSEFLGDAFLKPRTHNRVGDMLLEKEVVVTVYKVKDETDYNNKRVYIVLELKEE
RVLKANKNVTIIPKSSSYTKK"

/gene="TC0438"
/note="similar to GP:6013469; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="adherence factor"
/protein_id="AA039292.1"
/db_xref="GI:7190481"

Qy	10	taaaagtgaagcttcacaccgcatgaactaagcgcgaagcgaggctgcactiga	69
Db	9518	TAAGGAGATCGCCTCTTAATAATTCGAGATAGGATAGAGCTGGTTGAGACAGA	9577
Qy	70	ctggtgaagttcacactgaagtgactactatctcaattatcacca	117
Db	9578	CTGGCAAAATTCGAGATAGATAGCTTACGATTACGATTTCCTCGAC	9625

Query Match

Best Local Similarity

Matches

60; Conservative

0; Mismatches

48; Indels

0; Gaps

25.4%; Score 31.2; DB 1; Length 23366;

55.6%; Pred. No. 4;

RESULT 19
CEK10D6 30494 bp DNA INV 25-OCT-2000
LOCUS Caenorhabditis elegans cosmid K10D6, complete sequence.
DEFINITION
ACCESSION 274040
VERSION 274040.1 GI:1403245
KEYWORDS HTG; GABA receptor.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
1 (bases 1 to 30494)
none.
REFERENCE
AUTHORS Genome sequence of the nematode C. elegans: a platform for
TITLE Investigating biology. The C. elegans Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613
REMARK The C.elegans Sequencing Consortium.
Erratum: [[published errata appear in Science 1999 Jan
1:283(5398):35 and 1999 Mar 26:283(5410):2103 and 1999 Sep
3:285(5433):1493]]
2 (bases 1 to 30494)
White,S.
REFERENCE
AUTHORS Direct Submission
TITLE Submitted (14-JUN-1996) Nematode Sequencing Project, Sanger Centre,
JOURNAL Hinxton, Cambridge CB10 1RO, England and Department of Genetics,
Washington University, St. Louis, MO 63110, USA. E-mail:
jes@sanger.ac.uk or twenematode.wustl.edu
Coding sequences below are predicted from computer analysis, using
Predictions from GeneFinder (P. Green, U. Washington), and other
available information.
Current sequence finishing criteria for the C. elegans genome
sequencing consortium are that all bases are either sequenced
unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is not the entire sequence of clone K10D6.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone K10D6 is at 1 in this sequence. The true
left end of clone T19B10 is at 30388 in this sequence. The start of
this sequence (1..108) overlaps with the end of sequence AL021173.
The end of this sequence (30388..30494) overlaps with the start of
sequence Z74043.
For a graphical representation of this sequence and its analysis
see: [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=K10D6)
name=K10D6.
FEATURES
source Location/Qualifiers
1..30494
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="V"
/clone="K10D6"
join(1925..2071,3486..3705,3814..3893,3953..4045,
4098..4174,4480..4528,4580..4905,4954..5404,5462..5539)
/gene="K10D6.1"
join(1925..2071,3486..3705,3814..3893,3953..4045,
4098..4174,4480..4528,4580..4905,4954..5404,5462..5539)
/gene="K10D6.1"
/note="predicted using GeneFinder
Similarity to Drosophila GABA receptor beta subunit
(SW:GAB DROME), contains similarity to Pfam domain:
PF00065 (Neurotransmitter-gated ion-channel), Score=148.6,
E-value=5.7e-43, N=2"
/codon_start=1
/protein_id="CA98509.1"
/db_xref="GI:3878477"
/translation="MGIQFLNLSISFPVYISFIFQLSMFLFPMKSGFIYHDY
MSSQHNKRTNSFTIFLLISLIDINCEIITIDSYSSDDPTTQKRNCTR
QPAIDKLNGTGYNKRIFIPENGVQVQVEFWQATISINETINDEMDIYINEMLD
PALKEHLNPKONLSVSHOVLEQIWTPNSCFVNSKFAEIHDSPFKNVFLMIYPNCTV

WNYRVNKGCDLSLELFLDIOECHLIYESFNYNQOYRMKRNKSAPSVNTKI
RLDPFELIKIESTRIISAPYAGMDEIHLVFEIRIYTWQAYLPTYLITFISMS
FSLGTAMPARTMGLVNAIMIEFGNIRNRPVSYVAIDVMVIMVSTFELSIL
ELIAIVGYKTKNEEGSKKCPCHKILDLNFEPSGLICREKRPALPERBSRANGIIR
FQYQDFWMSPEKIDRVSAIMPACAFINIIYMSYYIKKLEKAEAMENDRLL"
complement(join(6124..6262,6777..6830,6977..7187,
7236..7461,7969..8202))
/gene="K10D6.2b"
complement(join(6124..6262,6777..6830,6977..7187,
7236..7461,7969..8202))
/note="cdna EST yk28494.3 comes from this gene
cdna EST yk2554.5 comes from this gene
cdna EST yk525b5.3 comes from this gene"
/codon_start=1
/protein_id="CA98513.1"
/db_xref="GI:3878481"
/db_xref="SPRMBL:045669"
/translation="MGSIVQKLLALISVILVIGICITSAGCSPAMQVVDIREPRA
EHQHGIMDCIRAEKHVSVGDEYDEPLHCWKFPNSAELVIONTLNIDEGAGE
SEHHRFAMWKALIFLITSEFLAFISICGVCAPCPSTAPAFSISLFAMCSILA
DGIFFLAANRVDRFVQGWGTYEQRIGAFYLIHMGTLICIGAFVCTLLTTPKPYTG
DEEDGSKENLFTWQOQTPVTLVSNELPYRREPLLHDKFVSSQAGYRNPVPYRT
TSITQYRETSA"
complement(join(6150..6262,6410..6594,6777..6830,
6977..7187,7236..7461,7969..8202))
/gene="K10D6.2a"
complement(join(6150..6262,6410..6594,6777..6830,
6977..7187,7236..7461,7969..8202))
/note="cdna EST yk2554.3 comes from this gene
cdna EST yk2554.5 comes from this gene
cdna EST yk526b2.3 comes from this gene"
/codon_start=1
/protein_id="CA98514.2"
/db_xref="GI:6425186"
/db_xref="SPRMBL:045670"
/translation="MGSIVQKLLALISVILVIGICITSAGCSPAMQVVDIREPRA
EHQHGIMDCIRAEKHVSVGDEYDEPLHCWKFPNSAELVIONTLNIDEGAGE
SEHHRFAMWKALIFLITSEFLAFISICGVCAPCPSTAPAFSISLFAMCSILA
DGIFFLAANRVDRFVQGWGTYEQRIGAFYLIHMGTLICIGAFVCTLLTTPKPYTG
DEEDGSKENLFTWQOQTPVTLVSNELPYRREPLLHDKFVSSQAGYRNPVPYRT
TSITQYRETSA"
complement(join(6262,6406..6594,6977..7187,7236..7461,
7969..8202))
/gene="K10D6.2c"
complement(join(6262,6406..6594,6977..7187,7236..7461,
7969..8202))
/gene="K10D6.2c"
/codon_start=1
/protein_id="CAB54276.2"
/db_xref="GI:6425185"
/translation="MGSIVQKLLALISVILVIGICITSAGCSPAMQVVDIREPRA
EHQHGIMDCIRAEKHVSVGDEYDEPLHCWKFPNSAELVIONTLNIDEGAGE
SEHHRFAMWKALIFLITSEFLAFISICGVCAPCPSTAPAFSISLFAMCSILA
DGIFFLAANRVDRFVQGWGTYEQRIGAFYLIHMGTLICIGAFVCTLLTTPKPYTG
DEEDGSKENLFTWQOQTPVTLVSNELPYRREPLLHDKFVSSQAGYRNPVPYRT
TSITQYRETSA"
complement(join(13294..13356,13407..13595)
join(13294..13356,13407..13595)
/gene="K10D6.3"
/note="predicted using GeneFinder"
/codon_start=1
/protein_id="CA98510.1"
/db_xref="GI:3878478"
/translation="MSTQPSHSSNDLSIGTARIEVSVPRKPKNSTTPVARKV
EMKSKESDVDAKLVKSSLDQMKLEKRAPPTCEKAA".27301,27401..27463,
27716..27902,27954..28159,28296..28568)
/gene="K10D6.4"


```

CDS
/note="T11A7.20"
complement(join(432..513,610..671,753..886,979..1282,
1353..1445,1532..1668,1807..1957,2194..2301,2379..1282,
2590..2725,2803..2944,3049..3182,3286..3468,3518..3622,
3683..3870,3971..4067,4178..4219,4329..4412,4721..4903,
5265..5298,5418..5596,5983..6561,6592..6684,6974..7043,
7131..7185,7641..7825,7904..8002,8302..8388,8924..8986,
9178..9284,9595..9707,9796..10004,10208..10297..
10381..10495,10630..10818,11403..11611,11937..12153,
12524..12606,12705..12873,12977..13035))
/gene="At2g41700"
/codon_start=1
/product="putative ABC transporter"
/protein_id="AAC02761.2"
/db_xref="GI:6598351"
/translation="MILLAVTRVDTTHPHASNDKRTVEVGKNSPPEVILK
LIAEGDFLAPDDETNMIDILSKPELRLVTKIEKDIIELEYTISAHGVCSE
VANCNSPKTKGAVFHEGPHLEDYSIRLNHTWAFGPNVNSIMDTNCPYINDLEM
INTIPTMOYSFSGFLTLOQVDSFTIFASQNNNDPLSHNSLSALREFLPWLSPS
VIRNVPPTREYDEDFOSIVKSMGLFELKSDKTLVTFEFLGLSALMSFMIST
PFTRAKTAAVGTITFLFAFPFYVNDSEGMVLKVVAALSPAFALGINSFADYE
RAHGLRMSNIWRASSGSEFFYCLIMLLDSTLYCALGIYIDKVLPRENGVYNNFI
PSKYGRKKNNLNQRIKPFETDMPADIEVNGDEPDPVESISLEMRQDELDKRIQ
VNLHRVYASRRGCCAVNSIQLTLYENQISLGLHNGAGKSTTISMLGLPTSGD
ALILGNSITNNDEIRKELGVCPOHDIIEPELVRHLEMPAVLKGVEGSLKSTVD
MAEEGLSDKINTLVNALSGLKRLSLGILIGNSKVIIDEPISGMDPYSMRLTWQ
LIKIKKGRITLITLTHSMDEAELDGRIGIMANGSLKCCGSSIFLKHGVCYTTITV
KSTPYVAVAHIVHRIHIPSACVSEYGNISFKPLASIPCFENMRETESCKNSDS
DYPGIOSYGIYITTLIEVFLRAGCNLDIEDKOEDFVSPDKSSIVCIGSNCKSMQ
PKLLASGNDGAVIITSVAKARLIVAAMVLIIGFTISIQCCGSSISREMRHCKAL
FIKRARSACDRKRYAFQFIIPAVFLFGLFLQICPHDQCSITLTVAFWPLSGK
GGGCPIDPELSYPAKEVAYIEGGMIQPLRMTSKFPRKALADADAAGTLCPT
LISMSEFLMSFDSQOSREGLSHSDSCNHNDGSLGAGYALIVNAGCOHAGPIYINMH
AAILATGCKNMKTIOIOTNRHPIPTKTOIRHODDAESALIVNAGSEFISFASV
LYKEKEVAKRHQOLISGVSYLSWISTYWDISTLPESTRHIIIFPAGLDFGIG
RFLPVLMLLEYGLAISSTYCLTFEFTSHMAQATSYSLVPLISLFEFSFSSNVL
WHPFSGILINVISFVGLIPATASNSYLKELIFRYALQNFRLSPQFCSDGAL
LALROGMKDSHGVFEWNVGASICYLSTVRIEYCYSMALLSFEFGIDTKLSLI
YIIGASRLTELIDRYSTFSTPELKNSTALSTDMEDDIDVOBERDRTVIGLSDN
IMLYLONLRKVPGRKHGPKYAVOSIPEFSVAGGFEIDNGAGKCTTILSMISSE
PPTSGTAFIFGCDIYASPKAIRHIGYCGPAPALPEYLIVKHELEIYAIKRVYDRI
DNVYTERKLEPDLKSHKSFSTLSGNNRKLISVAILAMIGDPVYILDPSYGMDBVA
KREMDVLSRLTSRSGKTAVILITTHSMNAQALCIRIGMVGRLGICSPHLKRY
GNHLELVPEYGVKPNESVNYELNFCQIIOQWLFNTPQRLSGLDEVCIGVSDS
ITPDFAASEILSPENVORIAKFLGNERNVSTLVPLDEEYREFDQSEDFRGG
IPLPIFAEMWLKEKFSALDSFLOSFPATKSCNGLSIKQLPRGEGSLADAFG
HLEHNRNRIGIAEYSOSTLETIFINHPANS"
complement(658..>83145)
/note="Sequence from clone T11A7"
/misc_feature
complement(13464..13494)
/rpt_family="POLY_A"
repeat_region
complement(join(<17467..17523,17579..18237,18321..18382,
18729..18779,18860..18933,19114..19202,19292..19300,
19447..19529,19870..>20096))
/gene="At2g41710"
complement(<17467..>20096)
/note="T11A7.19; PFAM domain (PF00847)"
/gene="At2g41710"
complement(join(17467..17523,17579..18237,18321..18382,
18729..18779,18860..18933,19114..19202,19292..19300,
19447..19529,19870..>20096))
/gene="At2g41710"
/codon_start=1
/product="putative AP2 domain transcription factor"
/protein_id="AAC02777.1"
/db_xref="GI:2887500"
/translation="MASVSSDQGRKTGACGGGGESSETVAAQDMILYRGEKKA
KKEGCTAKERTSKMPKAGRSSIYRVTYHRTMGREAHLMPSKSTNNQNNKKG
OYLYGAYDEBAARAYDLAAKMGPGLLNPDPYLDREEMGNLSREYLSLR
RKSQSGRGIAYRGIOASRMDASASMPPEYNSNHVAGACRQEGDLSFLER
KIDLIGYIKMMGANKRROPRESSKASEDANVDACTELTLEHTSHATEPYRAPNIGV
LCGTOKREKEISSPSSSALISQSPARKSLEEVYKLIQECSENNENDNANRNIIM
EKNGKAIEKRPVSHGVALGAAALSLQKSWPLSLTAPLITNNTPLADPLITM

```

```

mRNA
TFPLSGSSLTSEVRSHISIEPTSEVTKETSTYSLYPOEK"
complement(join(<20937..21001,21098..21218,21326..21430,
21547..21679,21776..21980,22070..22175,22312..23385,
23520..23625,23720..23892,23976..>24320))
/gene="At2g41720"
complement(<20937..>24320)
/note="T11A7.18"
/gene="At2g41720"
complement(join(20937..21001,21098..21218,21326..21430,
21547..21679,21776..21980,22070..22175,22312..23385,
23520..23625,23720..23892,23976..>24320))
/gene="At2g41720"
/codon_start=1
/product="putative salt-inducible protein"
/protein_id="AAC02776.1"
/db_xref="GI:2335106"
/translation="MATVTFNFKLVTPPESSRADKPGATASDAFOEKSVSUNYDGE
HEVSVNIGLRADIPRRYRIRENDRFQDMSVEVDRLMALNMEVDGLNSVY
GFRARNFPEVLIRELSRCCIIELCVAVFWMKIQNKYCARNDIYNKRIILHARHNV
OARGLEFEOKNSKPDAEYDALINAHGRAGQMRAMNIMDMRLRAALAPRSSTYNN
LINAGSSGNMRBEALEVCKMKMDNGVGPLVHNTVLSAYKSGROYSKALSYEELMG
AVRPPTTFENIIYCLSKGSSQALDIPNSRREKARACRPVYFTSMLHYSKG
ETENCAVFEAMABSLAPNIVSYNALCAVAVHSGTALSVLGDIKONGIIPDVS
YTCLINSYGRSROPKAKAEVFLMRERKRPVYTYNALIDAGSNGFLAAVEIFRQ
MEODGIKPNVVSCTLLAACSRSKRVNDVYLSAOSRGININTAAVSAIGSYTNO
AELEKAIAYOSMRKKRVKADSVYFTLISGSCRSKPYEAIKLEMBDSTPLKE
GOVTEAESIFENMKMGCPDVAATYFTLSMLAHNASMKWAKACFLFEMANGIEPSI
ACSALMRANRKGQSPNVYVNDIMMERKIPITGAFEFISFCNTLOMKRADILO
MDPPLTSLIGITNOMLHLFGSGVREEMMKFLKFTINSGVGINRTYAILLEHLLA
VNMRKRYIVLILFEMSGAGIOPSQWYRDIISFGREIVIFPEPLIRKLESLNNKGE
LIPTFHECTEL"
<26681..>27040
/gene="At2g41730"
<26681..>27040
/gene="At2g41730"
/note="T11A7.17; predicted by genefinder"
26681..27040
CDS
Query Match 24.4%; Score 30; DB 8; Length 83145;
Best local similarity 54.5%; Pred. No. 11;
Matches 60; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
QY 12 aatgtagatgccttaaaccttgatgacaaatgcgaatgcgagagtcgactgact 71
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72816 AATGAGAAATTCACATCGGCAAGATGACATGATGATTTAAGAAAGCTTCGTCAC 72757
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 72 ggttaagttcaatgaagtgaactatctcaattatccacaaca 121
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72756 TGTAAAGAGCAAAATGTGAGGTCACTCTTTCCTTTACAAACAA 72707
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 21
LOCUS AC027346/c 68575 bp DNA HTG 25-APR-2001
DEFINITION Homo sapiens chromosome 16 clone CTD-2519D12, WORKING DRAFT
ACCESSION AC027346
VERSION 2 (bases 1 to 68575)
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 68575)
AUTHORS DOE Joint Genome Institute.
JOURNAL Unpublished
TITLE Sequencing of Human Chromosome 16
2 (bases 1 to 68575)
DOE Joint Genome Institute.
DIRECT SUBMISSION Submitted (30-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Apr 25, 2001 this sequence version replaced gi:8576174.

```


Db 53767 TTATCTATAAATCTAGACATTAATTGATCTTAATTAGCCTTACT 53814

RESULT 25
AC016271/c
LOCUS AC016271 80065 bp DNA HTG 13-JUL-2000
DEFINITION Homo sapiens clone RP11-19F12, LOW-PASS SEQUENCE SAMPLING.
AC016271
ACCESSION AC016271.2 GI:9123891
VERSION
KEYWORDS HTGS_PHASED.
SOURCE
ORGANISM
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 80065)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
JOURNAL
Unpublished
2 (bases 1 to 80065)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Bouckgalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., D'Atellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Ferrera, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
Galaqan, J., Gardyna, S., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lehoczky, J., Lieu, C., Locke, K., Macdonald, P., Marquis, N.,
McEwan, P., McGuirk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrison, D., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Teisye, S., Tirrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
Direct Submission
Submitted (24-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jul 13, 2000 this sequence version replaced gi:6467063.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L3728
Center clone name: 19_F_12

* NOTE: This record contains 82 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 903: contig of 903 bp in length
* 904 1003: gap of 100 bp
* 1004 1895: contig of 892 bp in length
* 1896 1995: gap of 100 bp
* 1996 2890: contig of 895 bp in length
* 2891 2990: gap of 100 bp
* 2991 3816: contig of 826 bp in length
* 3817 3916: gap of 100 bp
* 3917 4776: contig of 860 bp in length
* 4777 4876: gap of 100 bp
* 4877 5746: contig of 870 bp in length
* 5747 5846: gap of 100 bp
* 5847 6741: contig of 895 bp in length

* 6742 6841: gap of 100 bp
* 6842 7705: contig of 864 bp in length
* 7706 7805: gap of 100 bp
* 7806 8688: contig of 883 bp in length
* 8689 8788: gap of 100 bp
* 8789 9660: contig of 872 bp in length
* 9661 9760: gap of 100 bp
* 9761 10661: contig of 901 bp in length
* 10662 10761: gap of 100 bp
* 10762 11625: contig of 864 bp in length
* 11626 11725: gap of 100 bp
* 11726 12619: contig of 894 bp in length
* 12620 12719: gap of 100 bp
* 12720 13621: contig of 902 bp in length
* 13622 13721: gap of 100 bp
* 13722 14562: contig of 861 bp in length
* 14563 14682: gap of 100 bp
* 14683 15537: contig of 855 bp in length
* 15538 15637: gap of 100 bp
* 15638 16517: contig of 880 bp in length
* 16518 16617: gap of 100 bp
* 16618 17504: contig of 887 bp in length
* 17505 17604: gap of 100 bp
* 17605 18479: contig of 875 bp in length
* 18480 18579: gap of 100 bp
* 18580 19458: contig of 879 bp in length
* 19459 19558: gap of 100 bp
* 19559 20428: contig of 870 bp in length
* 20429 20528: gap of 100 bp
* 20529 21424: contig of 896 bp in length
* 21425 21524: gap of 100 bp
* 21525 22423: contig of 899 bp in length
* 22424 22523: gap of 100 bp
* 22524 23367: contig of 844 bp in length
* 23368 23467: gap of 100 bp
* 23468 24369: contig of 902 bp in length
* 24370 24469: gap of 100 bp
* 24470 25324: contig of 855 bp in length
* 25325 25424: gap of 100 bp
* 25425 26267: contig of 843 bp in length
* 26268 26367: gap of 100 bp
* 26368 27235: contig of 868 bp in length
* 27236 27335: gap of 100 bp
* 27336 28210: contig of 875 bp in length
* 28211 28310: gap of 100 bp
* 28311 29200: contig of 890 bp in length
* 29201 29300: gap of 100 bp
* 29301 30174: contig of 874 bp in length
* 30175 30274: gap of 100 bp
* 30275 31162: contig of 888 bp in length
* 31163 31262: gap of 100 bp
* 31263 32105: contig of 843 bp in length
* 32106 32205: gap of 100 bp
* 32206 33111: contig of 906 bp in length
* 33112 33211: gap of 100 bp
* 33212 34119: contig of 908 bp in length
* 34120 34219: gap of 100 bp
* 34220 35103: contig of 884 bp in length
* 35104 35203: gap of 100 bp
* 35204 36064: contig of 861 bp in length
* 36065 36164: gap of 100 bp
* 36165 37033: contig of 869 bp in length
* 37034 37133: gap of 100 bp
* 37134 38021: contig of 888 bp in length
* 38022 38121: gap of 100 bp
* 38122 39018: contig of 897 bp in length
* 39019 39118: gap of 100 bp
* 39119 40034: contig of 916 bp in length
* 40035 40134: gap of 100 bp
* 40135 41047: contig of 913 bp in length
* 41048 41147: gap of 100 bp
* 41148 42068: contig of 921 bp in length
* 42069 42168: gap of 100 bp

```

* 42169 43048: contig of 880 bp in length
* 43049 43148: gap of 100 bp
* 43149 44027: contig of 879 bp in length
* 44028 44127: gap of 100 bp
* 44128 44989: contig of 862 bp in length
* 44990 45089: gap of 100 bp
* 45090 45956: contig of 867 bp in length
* 45957 46056: gap of 100 bp
* 46057 46940: contig of 884 bp in length
* 46941 47040: gap of 100 bp
* 47041 47939: contig of 899 bp in length
* 47940 48039: gap of 100 bp
* 48040 48830: contig of 791 bp in length
* 48831 48930: gap of 100 bp
* 48931 49793: contig of 863 bp in length
* 49794 49893: gap of 100 bp
* 49894 50799: contig of 906 bp in length
* 50800 50899: gap of 100 bp
* 50900 51768: contig of 869 bp in length
* 51769 51868: gap of 100 bp
* 51869 52723: contig of 855 bp in length
* 52724 52823: gap of 100 bp
* 52824 53686: contig of 863 bp in length
* 53687 53786: gap of 100 bp
* 53787 54671: contig of 885 bp in length
* 54672 54771: gap of 100 bp
* 54772 55656: contig of 885 bp in length
* 55657 55756: gap of 100 bp
* 55757 56644: contig of 888 bp in length
* 56645 56744: gap of 100 bp
* 56745 57637: contig of 893 bp in length
* 57638 57737: gap of 100 bp
* 57738 58616: contig of 879 bp in length
* 58617 58716: gap of 100 bp
* 58717 59555: contig of 839 bp in length
* 59556 59655: gap of 100 bp
* 59656 60555: contig of 900 bp in length
* 60556 60655: gap of 100 bp
* 60656 61541: contig of 886 bp in length
* 61542 61641: gap of 100 bp
* 61642 62547: contig of 906 bp in length
* 62548 62647: gap of 100 bp
* 62648 63491: contig of 844 bp in length
* 63492 63591: gap of 100 bp
* 63592 64444: contig of 853 bp in length
* 64445 64544: gap of 100 bp
* 64545 65420: contig of 876 bp in length
* 65421 65520: gap of 100 bp
* 65521 66396: contig of 876 bp in length
* 66397 66496: gap of 100 bp
* 66497 67376: contig of 880 bp in length
* 67377 67476: gap of 100 bp
* 67477 68377: contig of 901 bp in length
* 68378 68477: gap of 100 bp
* 68478 69360: contig of 883 bp in length
* 69361 69460: gap of 100 bp
* 69461 70336: contig of 876 bp in length
* 70337 70436: gap of 100 bp
* 70437 71288: contig of 852 bp in length

```

Query Match 24.1%; Score 29.6; DB 2; Length 80065;
Best Local Similarity 53.6%; Pred. No. 16;
Matches 59; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

```

Oy 3 ggtattgtaaatgtatagatgccttacaacccctgatgacctaatgagcgaatgcgaggct 62
Db 12614 GGGCTTCAAAAGTGAAGTATGGAACGTGGGAAGCAATGGCACACACAAAGATGN 12555
Oy 63 gcaactgactgtaagttcacatgtaagtgatctactcaattat 112
Db 12554 GNACTCTCTGGCGACACTGCACAGCAAGGAAACATGTATCTCAGATGTAT 12505

```

```

RESULT 26
AC005333 94797 bp DNA INV 29-JUL-1998
LOCUS AC005333
DEFINITION Drosophila melanogaster DNA sequence (Pls DS06477 (D116) and
DS03026 (D126)), complete sequence.
ACCESSION AC005333
AC004243 AC001697 AC001698 AC001699 AC001700 AC002577
AC003702 AC001701 AC001702 AC001703 AC001704 AC001705 AC003129
AC001706 AC001707 AC001708 AC001709 AC001710 AC001711 AC002578
AC001712 AC003703 AC003704 AC004271 AC003138 AC003716
AC005333.1 GI:3347820
VERSION
KEYWORDS
SOURCE
ORGANISM
Drosophila melanogaster (Subclones in sac from P1 clones DS06477
(D116) and DS03026 (D126)) DNA.
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 94797)
Celisner,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
Svitskas,R.R., Harris,N.L., Abdayani,A., Arcaina,T.T., Baxter,E.,
Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Paclebo,J.M., Park,S.,
Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.
Sequencing of Drosophila chromosome 2L, region 25F3-26A2
Unpublished (1997)
2 (bases 1 to 94797)
Celisner,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
Svitskas,R.R., Harris,N.L., Abdayani,A., Arcaina,T.T., Baxter,E.,
Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Paclebo,J.M., Park,S.,
Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.E.
Direct Submission
Submitted (29-JUL-1998) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (http://fruitfly.berkeley.edu/sequence/) or send
email to drosophila@mc.lbl.gov.
Library locations: 45-68, 50-32.
FEATURES
source
1..94797
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2L"
/map="25F3-26A2"
/clone="Pls DS06477 (D116) and DS03026 (D126)"
/note="DS06477 (d116) extends from bp 1 to bp 84,543 and
DS03026 (d126) is a bridge extending from bp 83,209 to bp
94,797."
BASE COUNT 27561 a 19822 c 19879 g 27535 t
ORIGIN
Query Match 24.1%; Score 29.6; DB 3; Length 94797;
Best Local Similarity 54.6%; Pred. No. 16;
Matches 59; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
Oy 6 atttgtaaatgtatagatgccttacaacccctgatgacctaatgagcgaatgcgaggctcga 65
Db 11285 ATTGTATATATTAATGATCGTTTCTTAACCTAATGGCAAGGTTTATTAAGAACTAATGAT 11344
Oy 66 ctgactgtaagttcacatgtaagtgatctactcaattat 113

```



```
TITLE      Direct Submission
JOURNAL    Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT    On Oct 9, 2000 this sequence version replaced gi:7297008.
FEATURES   location/Qualifiers
            1..259973
             /organism="Drosophila melanogaster"
             /db_xref="taxon:7227"
             /chromosome="2L"
             /join(<16590..16757,16815..17687,17749..>17886)
             /gene="CG7236"
             /product="CT22313"
             /db_xref="FLYBASE:FBan0007236"
             /db_xref="FLYBASE:FBgn0031730"
             /db_xref="FLYBASE:FBgn0031730"
             /map="25F4-26A1"
             /gene="CG7236"
             /db_xref="FLYBASE:FBan0007236"
             /db_xref="FLYBASE:FBgn0031730"
             /join(16590..16757,16815..17687,17749..17886)
             /gene="CG7236"
             /note="CG7236 gene product"
             /codon_start=1
             /db_xref="FLYBASE:FBan0007236"
             /db_xref="FLYBASE:FBgn0031730"
             /protein_id="AAF52279.1"
             /db_xref="GI:7297009"
             /translation="MDREKLSRLGSGYGVYKCRDRTGALVAVKRFEESEDDPAI
             RKIALREIRLKNLKNPVLSEVFRRRRLHVFECFLVLEHLEHHPGCEPHL
             TKOICYOTLLGAVCHGQGLCHRDIKPENILITLAGOVYKCDGFEARMISPEENTDY
             VATRWYRABELLVGDTOYGTPDVMAIGCLFAELVGEALMGRSDVDLYLIRRTLG
             DLPRHIOIFGONEYFGKIGITLPPPTLEPDERKMPKSSQNPULTIDELKCLDKDPTK
             RWSCERLKHSHFYDYIAKQRELEHVNLSLEANLQQOASQFMLATTAQOCLQSPA
             QAAALAAARDKSKTSNTSLPLPSTQHHNHPQDVKLQPLKNANLILHRTHHNPTI"
             complement(join(<19824..20287,20339..>20411))
             /gene="CG14007"
             /product="CT33563"
             /db_xref="FLYBASE:FBan0014007"
             /db_xref="FLYBASE:FBgn0031731"
             /evidence=not_experimental
             complement(<19824..>20411)
             /gene="CG14007"
             /map="25F5-26A1"
             /db_xref="FLYBASE:FBan0014007"
             /db_xref="FLYBASE:FBgn0031731"
             /evidence=not_experimental
             complement(join(19824..20287,20339..20411))
             /gene="CG14007"
             /note="CG14007 gene product"
             /codon_start=1
             /db_xref="FLYBASE:FBan0014007"
             /db_xref="FLYBASE:FBgn0031731"
             /evidence=not_experimental
             /protein_id="AAF52280.1"
             /db_xref="GI:7297010"
             /translation="MIMLKQINIGLLITLLGCFYLLFVGYNACTGAMFOIIGLV
             KSARCYELGDDPPEHPLAMRSNKDSEKPLVLEHMHMDHNDINDEKRLSLGLMPR
             ROPVSVLVPQAAVPOPPLKWMQLOALSTGDDHQDPDSVYRNWHTNELLNCGDVO
             RKSELNRYAENOLHRWY"
             complement(join(<22642..23763,23844..24295))
             /gene="CG11149"
             /product="CT31163"
             /db_xref="FLYBASE:FBan0011149"
             /db_xref="FLYBASE:FBgn0031732"
             complement(<22642..>24295)
             /gene="CG11149"
             /map="25F5-26A1"
             /db_xref="FLYBASE:FBan0011149"
             /db_xref="FLYBASE:FBgn0031732"
             complement(join(22734..23763,23844..24295))
             /gene="CG11149"
             /note="CG11149 gene product"

CDS
            /codon_start=1
            /db_xref="FLYBASE:FBan0011149"
            /db_xref="FLYBASE:FBgn0031732"
            /protein_id="AAF52281.1"
            /db_xref="GI:7297011"
            /translation="MATISFCNRKRYVRLMPVLLFNVALILWBNLQAOAKLASSNF
            GGQSVGLILSNESNPSSRSANFHLANQHKIITRESATTKELRSLCRDSSLNF
            ERLQGEFEMLONLVNKRKSRREVGCESVYTTTNDFTFFNNLEVVNSRMKAPVFAI
            HPPGDYDLNTLTDALIRYVRNCLPESDSIDWVSFHYFEPNQMPEYVPYDEAEVLMYH
            MCTLANGSLVSPLYTOIPTDYSKSRANLTYPINVRNIAIARLATNTHIFACDIELYP
            SGVFQDFLODMVARNHVSVALDPPRRRNVPLPVEIETGAKVDPDELLATYRKQO
            AOVFHLKLCPTCHTIPGQEWMLNTRSRADHGLHYFSKALRKRRFRAPFVSDNTEP
            LDEBVTMGSGSNRIQVYSCQVSYSDCLLONVAMCLDIEYHVLSPAFLYHSP
            GIKQSSKADSTRLQYAKETKFTIKKIIEPRVLFGKASAKT"
            join(25407..25577,25629..27907,27971..28124)
            /gene="slp1"
            /product="CT22315"
            /db_xref="FLYBASE:FBan0007238"
            /db_xref="FLYBASE:FBgn0024191"
            <25407..>28124
            /gene="slp1"
            /note="CG7238"
            /map="25F5-26A1"
            /db_xref="FLYBASE:FBan0007238"
            /db_xref="FLYBASE:FBgn0024191"
            join(25469..25577,25629..27907,27971..28102)
            /gene="slp1"
            /note="slp1 gene product"
            /codon_start=1
            /db_xref="FLYBASE:FBan0007238"
            /db_xref="FLYBASE:FBgn0024191"
            /protein_id="AAF52282.1"
            /db_xref="GI:7297012"
            /translation="MSDNDYERFEFTDYDLDFENININRPGRSRHOQIYIWNADSE
            EESGEGGCKRRGRARAKPDYTMVNVNAGIGIOAGKKKKALQADDEGSGQKEAY
            ADGGEEDSASAGRPARGDPPSSNSSEEPPTLSRKOPSTTPDHRSHIASERV
            CANEHTRGICAKLLIQMGYEPGKGLGNDLOGISHPVOAHYRKRGAIAGYGPETAS
            IGGTKNKSIRKVEDYREAKERFDQDNKRRKGSAGAEPMERQKRYTKSYVEETIAGS
            HTSGILSEKLSKGLGVNVIDMTGPEKRSVGHALGOAKITPEETLYDEATEIAGS
            APACVFAMPELTHNLQILVISOEQOIIAIDQNEEGSSQQALESERKLEIETOLER
            NHITRLEESLEVERLIDNPDLSPQARLRELLVYAAEFHFGIADLAAGVIAVL
            LKRELVOQPLENTEPEPLIKKRWGMQGDAAEQOPRNVFDPSYLIWAGVMSPR
            SSAANQREHPNPAASLIDNAPLILPSVYDLSYEQLYVRLVAGVQEMPPLDTYVI
            DSWLPMHAILIGSKLEAVYPOIRSKGLIALAVSPDRSRAVLTWQKAFPEEQO
            BFLQRYIVPKLOATLGLDLINPMQDLELQOVWEMHELIDPMYMAOLLDRHFEPRM
            QVLYVWLNQSDPYAEISRMYTGMKSMLEPQVPSYKHEHLRALLEIMHRASTLLQP
            TYTPPPPPVPAPVIMMDLIHPPAOLEFKEVLSQOCADIGIIPAPLPGRRMGKQIY
            RVGKLCFYCIDRHVCMVSDGSRSPKPVSLNHLERSGTGIL"
            complement(<31796..>33085)
            /gene="CG14006"
            /product="CT33562"
            /db_xref="FLYBASE:FBan0014006"
            /db_xref="FLYBASE:FBgn0031733"
            /evidence=not_experimental
            complement(<31796..>33085)
            /gene="CG14006"
            /map="25F5-26A1"
            /db_xref="FLYBASE:FBan0014006"
            /db_xref="FLYBASE:FBgn0031733"
            /evidence=not_experimental
            complement(31796..33085)
            /gene="CG14006"
            /note="CG14006 gene product"
            /codon_start=1
            /db_xref="FLYBASE:FBan0014006"
            /db_xref="FLYBASE:FBgn0031733"
            /db_xref="FLYBASE:FBgn0031733"
            /protein_id="AAF52283.1"
            /db_xref="GI:7297013"
            /translation="MMAFLAFVILGLNTVAASSFVMDNMFVEEQATYLRPGEPHIOE
            GQTLINMLYVNNVOYVCOPTAFSPMNVPOSNRLRLQIAAGEDYTOYKQATYREVNH
            AHSQLDCEHEKPLGHSAADPRIVPAIYEHSCYGIYTSKRPVLTLEVTSFLERS
```

```

Query Match ..... 24.1%; Score 29.6; DB 3; Length 259973;
Best Local Similarity 54.6%; Pred. No. 17;
Matches 59; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

Oy 6 attgtaaatgtgagatgcttaccacccctgatagctaataatggcgaatgcgagggctgca 65
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 210464 ATTGATATATATATAGATCGCTTCTTAACCTAAGTACGACAGTTTAAATAGAACTAACTGAT 210405
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Oy 66 ctgactggtaagtttcacatgtaagtgatactatcttcataatttatt 113
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 210404 TTATCTCATTAACCTTAGCATCTAAATATGATTAATTAATGACACTTACT 210357
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 29
AC021110
LOCUS Homo sapiens chromosome 18 clone RP11-238P13, WORKING DRAFT
DEFINITION AC021110
AC021110
ACCESSION AC021110.3 GI:8439978
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 155172)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 155172)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JAN-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jun 10, 2000 this sequence version replaced gi:7021826.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0238P13
----- Summary Statistics -----
Sequencing vector: M13; 73%
Sequencing vector: plasmid; 27%
Chemistry: Dye-Primer ET; 73% of reads
Chemistry: Dye-terminator Big Dye; 27% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 151162 bases at least Q40
Consensus quality: 152443 bases at least Q20
Consensus quality: 153249 bases at least Q20
Insert size: 177000; agarose-fp
Insert size: 154672; sum-of-contrigs
Quality coverage: 4.08 in Q20 bases; agarose-fp
Quality coverage: 4.69 in Q20 bases; sum-of-contrigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 9265: contig of 9265 bp in length
*
* 9266 9365: gap of unknown length
*
* 9366 26504: contig of 17139 bp in length
*
* 26505 26604: gap of unknown length
*
* 26605 44300: contig of 17696 bp in length
*
* 44301 44400: gap of unknown length
*
* 44401 62250: contig of 17850 bp in length
*
* 62251 62250: gap of unknown length

```

FEATURES	*	62351	105744:	contig of 43394 bp in length
	*	105745	105844:	gap of unknown length
	*	105845	155172:	contig of 49328 bp in length.
SOURCE				Location/Qualifiers
			1..155172	
			/organism="Homo sapiens"	
			/db_xref="taxon:9606"	
			/chromosome="18"	
			/clone="RP11-238P13"	
			1..9265	
			/note="assembly_name:Contig1"	
			9366..26504	
			/note="assembly_name:Contig2"	
			26605..44300	
			/note="assembly_name:Contig3	
			clone_end:T7	
			vector_side:right"	
			44401..62250	
			/note="assembly_name:Contig4"	
			62351..105744	
			/note="assembly_name:Contig5"	
			105845..155172	
			/note="assembly_name:Contig6"	
			51543 a 28385 c 28043 g 46698 t	503 others
BASE COUNT				
ORIGIN				
Query Match		23.9%	Score 29.4;	DB 2; Length 155172;
Best Local Similarity		55.3%;	Pred. No. 19;	
Matches 57;	Conservative	0;	Mismatches	46; Indels 0; Gaps 0;
QY	18	agaagccctaacacocctatgacctaatgcygcgaatcgagggctgcactgtagtgaag	77	
Db	83677	AGGCGCTTTATGCACCTTCATTAGCTTATGAAGAATAATGAATGGCTGCAATATGTGTAAAT	83736	
QY	78	tttcaatgttaagggtgatcattcttcctaatttatccacaca	120	
Db	83737	TACAGTAGTAATCTGTGCTGGAACCTCTCTTCTATACACA	83779	
RESULT	30			
LOCUS	ALJ359922			
DEFINITION	Human DNA sequence from clone RP11-70L8 on chromosome 9, complete			
ACCESSION	ALJ359922			
VERSION	ALJ359922.10			
KEYWORDS	GI:12191425			
SOURCE	HNG.			
ORGANISM	human.			
REFERENCE	Homo sapiens			
AUTHORS	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;			
TITLE	Mammalia; Eutheraia; Primates; Catarrhini; Hominiidae; Homo.			
JOURNAL	1 (bases 1 to 168656)			
COMMENT	Laird,G. Direct Submissions Submitted (17-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquereyes@sanger.ac.uk Clone request: clonerquest@sanger.ac.uk On Jan 13, 2001 this sequence version replaced gi:11878000. During difference assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated repeat sequence elements. Where the sequence is ambiguous, there is an annotation using the "unsure" feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: EMBL; SWI; SWISSPROT; Tr.; TREMBL; WP.; WORMPEP; Information			

FEATURES	source
*	135792 135891: gap of 100 bp
*	135892 1428334: contig of 7043 bp in length
*	142935 143034: gap of 100 bp
*	143035 147799: contig of 4965 bp in length
*	148000 148099: gap of 100 bp
*	148100 151734: contig of 3635 bp in length
*	151735 151834: gap of 100 bp
*	151835 155576: contig of 3742 bp in length
*	155577 155676: gap of 100 bp
*	155677 1559193: contig of 3517 bp in length
*	159194 159293: gap of 100 bp
*	159294 161359: contig of 2066 bp in length
*	161360 161459: gap of 100 bp
*	161460 163515: contig of 2056 bp in length
*	163516 163615: gap of 100 bp
*	163616 166614: contig of 2999 bp in length
*	166615 166714: gap of 100 bp
*	166715 169056: contig of 2342 bp in length
*	169057 169156: gap of 100 bp
*	169157 171527: contig of 2371 bp in length
*	171528 171627: gap of 100 bp
*	171628 173348: contig of 1621 bp in length
*	173249 173348: gap of 100 bp
*	173349 174615: contig of 1467 bp in length
	Location/Qualifiers
	1..174615

```

/misc_feature /organism="Homo sapiens"  

/misc_feature /db_xref="taxon:9606"  

/misc_feature /chromosome="18"  

/misc_feature /map="18q23"  

/misc_feature /clone="RP11-635B11"  

misc_feature . . 58823  

/misc_feature /note="assembly_fragment"  

/misc_feature 58924 . 73243  

/misc_feature /note="assembly_fragment"  

/misc_feature 73344 . 83572  

/misc_feature /note="assembly_fragment"  

/misc_feature 83673 . 94547  

/misc_feature /note="assembly_fragment"  

/misc_feature 94648 . 104597  

/misc_feature /note="assembly_fragment"  

/misc_feature 104698 . 114347  

/misc_feature /note="assembly_fragment"  

/misc_feature 114448 . 122050  

/misc_feature /note="assembly_fragment"  

/misc_feature 122181 . 128465  

/misc_feature /note="assembly_fragment"  

/misc_feature 128566 . 135761  

/misc_feature /note="assembly_fragment"  

/misc_feature 135892 . 142934  

/misc_feature /note="assembly_fragment"  

/misc_feature 143035 . 147959  

/misc_feature /note="assembly_fragment"  

/misc_feature 148100 . 151734  

/misc_feature /note="assembly_fragment"  

/misc_feature 151835 . 155576  

/misc_feature /note="assembly_fragment"  

/misc_feature 155677 . 159153  

/misc_feature /note="assembly_fragment"  

/misc_feature 159294 . 161359  

/misc_feature /note="assembly_fragment clone_end:T7 vector_side:right"  

/misc_feature 161460 . 163515  

/misc_feature /note="assembly_fragment clone_end:SP6 vector_side:left"  

/misc_feature 163616 . 166614  

/misc_feature /note="assembly_fragment"  

/misc_feature 166715 . 169056  

/misc_feature /note="assembly_fragment"  

/misc_feature 169157 . 171527  

/misc_feature /note="assembly_fragment"  

/misc_feature 171628 . 173248  

/misc_feature /note="assembly_fragment"  

/misc_feature 173349 . 174815  

/misc_feature /note="assembly_fragment"

```

BASE COUNT	52310	a	30923	c	32183	g	57399	t	2000	others
ORIGIN										
Query Match			23.9%	Score 29.4;	DB 2;	Length 174815;				
Best Local Similarity			55.3%;	Pred. No. 19;						
Matches	57;	Conservative	0;	Mismatches	46;	Indels	0;	Gaps	0;	
Oy	18	agatgccttacacacccttatgacctaatgycgaattgcgaggctgcactcgtctaag	77							
Db	71924	AGTGCTTTATTCACACTTATTAGCTTAAGAATAAATTGAATGCTGCACATATTTGTAAAT	71983							
Oy	78	tttcacatgtaagtgatcactatcttcaattatcttcacaca	120							
Db	71984	TACACTGTATATCTGTGGCTCAACCTCTTTTATACTAGACA	72026							

RESULT	32
LOCUS	AC091589
DEFINITION	AC091589 178056 bp DNA Homo sapiens chromosome 18 clone RP11-635B11 map 18, *** SEQUENCING IN PROGRESS ***, 2 ordered pieces.
ACCESSION	AC091589
VERSIONS	AC091589.6 GI:15022066
KEYWORDS	HTG; HTGS_PHAZE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE	human.
ORGANISM	Homo sapiens

REFERENCE	AUTHORS	TITLE	JOURNAL
1 (bases 1 to 178056)	Bliren, B., Linton, L., Nusbaum, C. and Lander, E.	Human sapiens chromosome 18, clone RP11-635B11	Unpublished
2 (bases 1 to 178056)	Bliren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S., ..		

AUTHORS

Bitren, B., Linton, L., Nushbaum, C., Lander, E., Allen, N., Anderson, S., Barua, N., Bastien, V., Boguslavsky, L., Boukhalter, B., Brown, A., Camarata, J., Campoliano, A., Chang, J., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Deakellano, K., Dewar, K., Diaz, O. S., Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S., Glnde, S., Goyette, M., Graham, L., Grand-Plerrie, N., Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Karatas, A., Larroque, K., Lamazares, R., Landers, R., Lehoczy, J., Levine, R., Liu, G., McLean, C., Macdonald, P., Marquis, N., Matthews, C., McCarthy, M., McMan, P., McKernan, K., McPheters, R., Meldrim, J., Menus, L., Minova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nord, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhkhang, P., Plerrie, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosettli, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Sougnez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Travers, M., Travis, N., Triggillo, V., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zembek, L., Zimmer, A., and Zody, M.

TITLE Direct Submission
 JOURNAL Submitted (07-MAY-2001) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 COMMENT On Jul 26, 2001 this sequence version replaced gi:14994291.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center

```

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: LI3182
Center clone name: 635_B_11

```


Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>
VECTOR: pcypac2
IMPORTANT: This sequence is not the entire insert of clone RPl-95121. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true right end of clone RPl-95121 is at 120515 in this sequence. The true left end of clone RPl-95121 is at 4877 in this sequence. The true right end of clone RPl-242N11 is at 100 in this sequence.

FEATURES

```
source
location/Qualifiers
1..120515
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone_1ib="RPl-95121"
/clone_1ib="RPCT-1"
1..62
repeat_region
/note="31 copies 2 mer at 71% conserved"
complement(640..1130)
misc_feature
/note="match: GSS: Em:AQ595695"
751..989
repeat_region
/note="AluIo repeat: matches 59..297 of consensus"
1620..2460
repeat_region
/note="LIM4 repeat: matches 4616..5481 of consensus"
2472..3677
repeat_region
/note="LIMC repeat: matches 1047..2247 of consensus"
3751..3929
repeat_region
/note="AluSg/x repeat: matches 134..310 of consensus"
3332..4089
repeat_region
/note="LIMC repeat: matches 890..1062 of consensus"
4090..4469
misc_feature
/note="match: GSS: Em:AQ013144"
4570..4746
repeat_region
/note="MT1A2 repeat: matches 200..374 of consensus"
4824..4931
repeat_region
/note="MT1A2 repeat: matches 55..162 of consensus"
5001..5050
repeat_region
/note="25 copies 2 mer aa 72% conserved"
complement(5063..5482)
misc_feature
/note="match: GSS: Em:AQ511994"
5279..5376
repeat_region
/note="L2 repeat: matches 2575..2678 of consensus"
5507..5951
misc_feature
/note="match: GSS: Em:AQ526871"
5515..6122
misc_feature
/note="match: GSS: Em:AQ530256"
complement(5957..6576)
misc_feature
/note="match: GSS: Em:AQ630441"
complement(6082..6597)
misc_feature
/note="match: GSS: Em:AQ693789"
6095..6259
repeat_region
/note="MIR repeat: matches 5..180 of consensus"
6362..6475
repeat_region
/note="MER5A repeat: matches 50..161 of consensus"
6591..7153
misc_feature
/note="match: GSS: Em:AQ379073"
6786..6934
repeat_region
/note="MER5B repeat: matches 1..173 of consensus"
7319..7428
repeat_region
/note="L2 repeat: matches 2575..2702 of consensus"
7609..7640
repeat_region
/note="16 copies 2 mer ac 87% conserved"
complement(10277..10730)
misc_feature
/note="match: GSS: Em:AQ839431"
10672..11208
misc_feature
/note="match: GSS: Em:AQ716569"
10853..11289
misc_feature
/note="match: GSS: Em:AQ082251"
12276..12566
repeat_region
/note="AluIo repeat: matches 1..298 of consensus"
12890..13106
repeat_region
/note="MIR repeat: matches 33..232 of consensus"
```

```
misc_feature
complement(12939..13175)
/note="match: GSS: Em:B85743"
13214..13514
misc_feature
/note="match: GSS: Em:AQ429929"
13216..13542
misc_feature
/note="match: GSS: Em:AQ835474"
13516..13551
repeat_region
/note="18 copies 2 mer tg 100% conserved"
13518..13695
misc_feature
/note="match: GSS: Em:AQ429929"
14150..14378
repeat_region
/note="MIR repeat: matches 14..241 of consensus"
15637..15675
repeat_region
/note="MADE1 repeat: matches 1..39 of consensus"
15901..16030
repeat_region
/note="MER5A repeat: matches 9..182 of consensus"
16173..16202
repeat_region
/note="15 copies 2 mer gt 96% conserved"
17819..18027
repeat_region
/note="MIR repeat: matches 7..225 of consensus"
18516..18938
repeat_region
/note="MT2CB repeat: matches 1..450 of consensus"
18934..19450
repeat_region
/note="MT2D repeat: matches 1..504 of consensus"
19465..19521
repeat_region
/note="MT2D repeat: matches 499..553 of consensus"
20072..20270
repeat_region
/note="LIMB5 repeat: matches 5936..6171 of consensus"
20444..24195
repeat_region
/note="LIM4 repeat: matches 2448..6279 of consensus"
25276..25640
repeat_region
/note="MT1A1 repeat: matches 1..365 of consensus"
26400..26653
repeat_region
/note="AluIo repeat: matches 1..287 of consensus"
30430..30896
repeat_region
/note="L1PA11 repeat: matches 5694..6165 of consensus"
33041..33473
repeat_region
/note="MT2FB repeat: matches 1..414 of consensus"
complement(34176..34667)
misc_feature
/note="match: GSS: Em:AQ111813"
34344..34474
repeat_region
/note="LIMC4 repeat: matches 7851..7977 of consensus"
34475..34502
repeat_region
/note="14 copies 2 mer ct 92% conserved"
34504..34733
repeat_region
/note="LIMC4 repeat: matches 7625..7857 of consensus"
35199..35312
repeat_region
/note="57 copies 2 mer tt 62% conserved"
35336..35696
misc_feature
/note="NAL1 repeat: matches 1170..1526 of consensus"
35971..36416
repeat_region
/note="match: GSS: Em:AQ152633"
36011..36113
repeat_region
/note="L1PA15 repeat: matches 1910..2012 of consensus"
36115..36503
repeat_region
/note="MER5B repeat: matches 1..399 of consensus"
36513..38620
repeat_region
/note="MER57-internal repeat: matches 1..2125 of consensus"
38624..38694
repeat_region
/note="AluJ/FRAM repeat: matches 228..295 of consensus"
38697..38999
repeat_region
/note="AluSx repeat: matches 1..300 of consensus"
39000..39345
repeat_region
/note="MER57-internal repeat: matches 2129..2596 of consensus"
39339..43921
repeat_region
/note="MER57-internal repeat: matches 2874..7531 of consensus"
43927..44325
repeat_region
/note="MER57B repeat: matches 1..403 of consensus"
44326..46056
repeat_region
/note="L1PA15 repeat: matches 2010..3753 of consensus"
```

```

repeat_region 46057..46329
/note="AluSp repeat: matches 17..290 of consensus"
repeat_region 46330..48228
/note="L1PA15 repeat: matches 3753..6155 of consensus"
repeat_region 48229..48888
/note="L1PA15 repeat: matches 5499..6153 of consensus"
repeat_region 49295..49393
/note="MIR repeat: matches 53..150 of consensus"
repeat_region 49721..49980
/note="L1M4 repeat: matches 2110..2375 of consensus"
misc_feature complement(50308..50438)
/note="match: GSS: Em: A0894428"
misc_feature complement(50343..50444)
/note="match: STS: Em: L17929"
repeat_region 51041..51106

Query Match 23.7%: Score 29.2; DB 9; Length 120515;
Best Local Similarity 57.8%; Pred. No. 22;
Matches 52; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

Qy 18 agatgcctaacacccctgactgaatgacgaatgagggctgactgactgtaag 77
Db 5308 AGAGGCATTCACAAACAAATACCTGCTCTCTGCGAGAAACACAGCTGTGAG 5367
Qy 78 ttacacgtgaagtgaactatcttcaa 107
Db 5368 AGAGGAAAAAGAAATACATATTACAA 5397

RESULT 35
AC013621/c AC013621 159073 bp DNA HTG 20-SEP-2000
LOCUS Homo sapiens clone RP11-11123, WORKING DRAFT SEQUENCE, 5 unordered
pieces.
AC013621 AC013621.5 GI:10198425
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 159073)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
JOURNAL Homo sapiens, clone RP11-11123
TITLE 2 (bases 1 to 159073)
REFERENCE Unpublished
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barna, N., Beckerly, R., Boguslavsky, L., Boukhalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., Dearellano, K., Dewar, K., Domino, M., Donelan, L., Doyle, M.,
Fitzgerald, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
Galligan, J., Gardyna, S., Grant, G., Hagos, B., Heath, A., Horton, L.,
Howland, J. C., Johnson, R., Jones, C., Kann, L., Karacas, A., Klein, J.,
Lehoczky, J., Liu, C., Locke, K., McDonald, P., Marquis, N.,
McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Traflet, S., Turrell, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W. J., Zimmer, A. and Zody, W.
COMMENT Direct Submission
Submitted (13-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Sep 20, 2000 this sequence version replaced gi:9123835.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information

```

```

Center project name: L3173
Center clone name: 11_123
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap, version 0.960731
Consensus quality: 156407 bases at least Q40
Consensus quality: 158013 bases at least Q30
Consensus quality: 158490 bases at least Q20
Insert size: 157000; agarose-fp
Insert size: 158673; sum-of-ctgigs
Quality coverage: 5.6 in Q20 bases; agarose-fp
Quality coverage: 5.6 in Q20 bases; sum-of-ctgigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 30588: contig of 30588 bp in length
* 30589 30688: gap of 100 bp
* 30689 82042: contig of 51354 bp in length
* 82043 82142: gap of 100 bp
* 82143 108087: contig of 25945 bp in length
* 108088 108187: gap of 100 bp
* 108188 134533: contig of 26346 bp in length
* 134534 134633: gap of 100 bp
* 134634 159073: contig of 24440 bp in length.
Location/Qualifiers
1..159073
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-11123"
/clone_lib="RPC1-11 Human Male BAC"
1..30588
/note="assembly-fragment"
clone_end:SP6
vector_side:left"
misc_feature 30689..82042
/note="assembly-fragment"
misc_feature 82143..108087
/note="assembly-fragment"
misc_feature 108188..134533
/note="assembly-fragment"
misc_feature 134634..159073
/note="assembly-fragment"
clone_end:T7
vector_side:right"
BASE COUNT 52902 a 27848 c 26048 g 51874 t 401 others
ORIGIN
Qy 10 taatgtgagatgccttaacacccgtgactgaatgacgaatgagggctgcacga 69
Db 4206 TAAATATCATGATATGACCCCTCTTCATCTTCATTTATTTAGTGTCTGACCTTATGTC 4147
Qy 70 ctggtgaattcactgaatgatacttactcaatctacacacaca 123
Db 4146 CTGGTATCTATCTTGAATAATTATATAGACTTCAACCTAGTCACATGAAA 4093

RESULT 36
AC022863 AC022863 159150 bp DNA HTG 17-MAR-2000
LOCUS Homo sapiens chromosome 4 clone RP11-352124 map 4, WORKING DRAFT
DEFINITION SEQUENCE, 5 unordered pieces.

```

ACCESSION ACO22863
VERSION ACO22863.3
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 159150)
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
JOURNAL Homo sapiens chromosome 4, clone RP11-352124
REFERENCE 2 (bases 1 to 159150)
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beckert, R., Beda, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A., Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeRellano, K., Dewar, K., Domino, M., Doyle, M., Feneator, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Gardina, S., Grant, G., Hagos, B., Heatford, A., Horton, L., Howland, D. C., Johnson, R., Jones, C., Kam, L., Karatas, A., Klein, J., Landers, T., Lehoczy, J., Levine, R., Lien, C., Liu, G., Locke, K., MacDonald, P., Marquis, N., McEwan, P., McGurk, A., McKenna, K., McPheters, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K., Plette, N., Pitsani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D., Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J., Zimmer, A. and Zody, M.
TITLE Direct Submission
JOURNAL Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Mar 11, 2000 this sequence version replaced gi:6980325. All repeats were identified using RepeatMasker:
Smith, A. F. A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L6194
Center clone name: 352_I_24
----- Summary Statistics
Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 155967 bases at least Q40
Consensus quality: 157848 bases at least Q30
Consensus quality: 158421 bases at least Q20
Insert size: 157000; agarose-1p
Insert size: 158750; sum-of-contigs
Quality coverage: 6.6 in Q20 bases; agarose-1p
Quality coverage: 6.5 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 3449: contig of 3449 bp in length
* 3450 3549: gap of 100 bp
* 3550 11150: contig of 7601 bp in length
* 11151 11250: gap of 100 bp
* 11251 48389: contig of 37139 bp in length
* 48390 48489: gap of 100 bp
* 48490 102960: contig of 54471 bp in length
* 102961 103060: gap of 100 bp
* 103061 159150: contig of 56090 bp in length.

FEATURES					
source		Location/Qualifiers			
		1..159150			
		/organism="Homo sapiens"			
		/db_xref="taxon:9606"			
		/chromosome="4"			
		/map="4"			
		/clone="RP11-352124"			
		/clone_lib="RPC1-11 Human Male BAC"			
		1..3449			
		/note="assembly_fragment"			
		3550..11150			
		/note="assembly_fragment"			
		clone_end:SP6			
		vector_side:left"			
		11251..48389			
		/note="assembly_fragment"			
		clone_end:F7			
		vector_side:left"			
		48480..102960			
		/note="assembly_fragment"			
		103061..159150			
		/note="assembly_fragment"			
BASE COUNT	50579 a	29976 c	29236 g	48959 t	400 others
ORIGIN					
Query Match		23.7%;	Score 29.2;	DB 2;	Length 159150;
Best Local Similarity	56.1%;	Pred.	No.23;		
Matches	55;	Conservative	0;	Mismatches	43; Indels 0; Gaps 0
QY	26	tacaacccgtgatacctaattggcgcaatgcgaagctgcactgactgtaagtccaat	85		
Db	13716	TAAACAATTCCTTTTAATAATCATGCAAGACAGAGCGTGTTAACCTGTAACCTA	13775		
QY	86	gtagtgatactacttccaatttatccacaacaca	123		
Db	13776	CTATCGTAGATANGCAGCTTCTATTATCGCGGTTAAC	13813		
RESULT	37				
LOCUS	AC021559				
DEFINITION	Homo sapiens chromosome 4 clone RP11-753L19 map 4,	HTG	23-APR-2000		
SEQUENCE	10 unordered pieces.				
ACCESSION	AC021559				
VERSION	AC021559.3	GI:7637771			
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.				
SOURCE	human.				
ORGANISM	Homo sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
REFERENCE	Birten,B., Linton,L., Nusbaum,C. and Lander,E.				
AUTHORS	1 (bases 1 to 170022)				
TITLE	Unpublished				
JOURNAL	2 (bases 1 to 170022)				
REFERENCE	Birten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barua,N., Beckerly,R., Beda,F., Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G., Castle,A., Choepel,T., Colangelo,M., Collins,S., Collimore,A., Cooke,P., DeRetillan,K., Dewar,K., Domino,M., Doyle,M., Feneslor,J., Ferreira,S., FitzHugh,W., Forrest,C., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kamp,L., Karatas,A., Klein,J., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Margulis,N., McEwan,P., McCurt,A., McKernan,K., McChetters,R., Meldrim,J., Menous,L., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K., Pierre,N., Pisanic,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy A., Santos R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J., Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody M.				

REFERENCE
AUTHORS

2 (bases 1 to 181542)
Britten, B., Linton, L., Nussbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,
Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Castle, A.,
Chapel, Y., Colangelo, M., Collins, S., Collamore, A., Cooke, P.,
DeRellano, K., Dewar, K., Domino, M., Doyle, M., Feneator, J.,
Ferrelia, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardner, S., Grant, G., Hagos, B., Heatford, A., Horton, L.,
Howard, J. C., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Lander, T., Lechoczky, J., Levine, R., Liu, C., Liu, G., Locke, K.,
Macdonald, P., Margulis, N., McEwan, P., McGurk, A., McKernan, K.,
McDonald, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Oliver, T. M., Peterson, K.,
Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Testaf, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W. J.,
Zimmer, A. and Zody, M.

TITLE
JOURNAL

COMMENT

Direct Submission
Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 1, 2000 this sequence version replaced g1:6957798.
All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: 102_H15
Center clone name: 102_H15

----- Summary Statistics
Sequencing vector: M13: M77815: 100% of reads
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 166930 bases at least Q40
Consensus quality: 173654 bases at least Q30
Consensus quality: 176417 bases at least Q20
Insert size: 182000; agarose-fp
Insert size: 179342; sum-of-contigs
Quality coverage: 3.9 in Q20 bases; agarose-fp
Quality coverage: 3.9 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1096: contig of 1096 bp in length
* 1097 1196: gap of 100 bp
* 1197 3560: contig of 2364 bp in length
* 3561 3660: gap of 100 bp
* 3661 5983: contig of 2323 bp in length
* 5984 6083: gap of 100 bp
* 6084 8818: contig of 2735 bp in length
* 8819 8918: gap of 100 bp
* 8919 11436: contig of 2518 bp in length
* 11437 11536: gap of 100 bp
* 11537 14541: contig of 3005 bp in length
* 14542 14641: gap of 100 bp
* 14642 18297: contig of 3656 bp in length
* 18298 18397: gap of 100 bp
* 18398 22136: contig of 3739 bp in length
* 22137 22236: gap of 100 bp
* 22237 25665: contig of 3429 bp in length
* 25666 25765: gap of 100 bp
* 25766 30336: contig of 4571 bp in length
* 30337 30436: gap of 100 bp

FEATURES

source

30437 34701: contig of 4265 bp in length
* 34702 34801: gap of 100 bp
* 34802 38748: contig of 3947 bp in length
* 38749 38848: gap of 100 bp
* 38849 48303: contig of 9455 bp in length
* 48304 48403: gap of 100 bp
* 48404 55751: contig of 7348 bp in length
* 55752 55851: gap of 100 bp
* 55852 65044: contig of 9193 bp in length
* 65045 65144: gap of 100 bp
* 65145 74306: contig of 9162 bp in length
* 74307 74406: gap of 100 bp
* 74407 85102: contig of 10696 bp in length
* 85103 85202: gap of 100 bp
* 85203 95862: contig of 10660 bp in length
* 95863 95962: gap of 100 bp
* 95963 111886: contig of 15924 bp in length
* 111887 111986: gap of 100 bp
* 111887 127546: contig of 15560 bp in length
* 127547 127646: gap of 100 bp
* 127647 144833: contig of 17187 bp in length
* 144834 144933: gap of 100 bp
* 144934 162954: contig of 18021 bp in length
* 162955 163054: gap of 100 bp
* 163055 181542: contig of 18488 bp in length.
Location/Qualifiers
1. .181542
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-102H15"
/clone.lib="RPC1-11 Human Male BAC"
1. .1096
misc_feature
/note="assembly-fragment"
1197. .3560
/note="assembly-fragment"
3661. .5983
/note="assembly-fragment"
6084. .8818
/note="assembly-fragment"
8919. .11436
/note="assembly-fragment"
11537. .14541
/note="assembly-fragment"
14642. .18297
/note="assembly-fragment"
18398. .22136
/note="assembly-fragment"
22237. .25665
/note="assembly-fragment"
25766. .30336
/note="assembly-fragment"
30437. .34701
/note="assembly-fragment"
34802. .38748
/note="assembly-fragment"
clone_end:77
vector_side:right
38849. .48303
/note="assembly-fragment"
48404. .55751
/note="assembly-fragment"
55852. .65044
/note="assembly-fragment"
65145. .74306
/note="assembly-fragment"
clone_end:SP6
vector_side:left
74407. .85102
/note="assembly-fragment"
85203. .95862
/note="assembly-fragment"


```

misc-feature      95963. 111886
                  /note="assembly-fragment"
misc-feature      111987. 127546
                  /note="assembly-fragment"
misc-feature      127647. 144833
                  /note="assembly-fragment"
misc-feature      144934. 162954
                  /note="assembly-fragment"
misc-feature      163055. 181542
                  /note="assembly-fragment"
BASE COUNT      60164 a 31041 c 31307 g 56823 t 2207 others
ORIGIN

```

```

Query Match      23.7%; Score 29.2; DB 2; Length 181542;
Best Local Similarity 53.5%; Pred. No. 23;
Matches 61; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 10 taaatgagatgacctaacacccatgacctaatgagcgaatgagagctgacatga 69
      ||||| || || || || || || || || || || || || || || || || ||
Db 58856 TAAATATCATGATATGACCCCTCTTCATCTTCAATTAATGCTGAGTTATGTC 58915

QY 70 ctggtgaagttcacatgtagtgatactatcttcaattatccacacacaa 123
      ||||| || || || || || || || || || || || || || || || || ||
Db 58916 CTGGTATCTATCTTGAATAATTATTAATGACTTCAACCTAGTCCACATGAAA 58969

```

```

RESULT 40
VAPPCGENE/c
LOCUS VAPPCGENE 1086 bp mRNA PLN 31-MAY-1997
DEFINITION V.aphylla mRNA for phosphoenolpyruvate carboxylase.
ACCESSION X91634
VERSION X91634.1 GI:2145478
KEYWORDS phosphoenolpyruvate carboxylase; ppc gene.
SOURCE Vanilla aphylla.
ORGANISM Vanilla aphylla.
          Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
          Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Orchidaceae;
          Vanilloid clade; Vanillineae; Vanilla.
REFERENCE 1 (bases 1 to 1086)
AUTHORS Gehrig,H.H.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1086)
AUTHORS Gehrig,H.H.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1995) H.H. Gehrig, TH-Darmstadt, Botanik FB10,
          Schmittsplanstr. 10, 64287 Darmstadt, FRG
FEATURES
source
1..1086
location/Qualifiers
/organism="Vanilla aphylla"
/db_xref="taxon:53061"
/tissue_type="shoot"
1..1086
/gene="ppc"
/gene="ppc"
<1..>1086
/gene="ppc"
/EC_number="4.1.1.31"
/codon_start=1
/product="phosphoenolpyruvate carboxylase"
/protein_id="CAA62828.1"
/db_xref="GI:2145479"
/translaton="SDSGKDAGRSLAASPYYKTOEELIVAKQGVKTLMEFGSGTV
GRGSGPTHLGILISOPPTILISLRYTVGEVIEQCFGERLCLFRLORTVATLEHGM
HHHISPKPEKMSLMDEMAVIAATKEKRSIVFQDPREVEYFRLATPELEYGRMKISRA
KKRPSGGIESLNAIPWAFWATRHPLVWGFGAFAFHVITKDIRNHLIEMYNAGH
PEFFVATIDLIENVEFKGNIGIASLIDELVSNLISFEMELRANYEETNRKLLIOVAGH
KDLLEGDHLKORLRDPDPYITTLNVCAYTLKALIRDQVTTQWNRHVSNETESRSAA
ELVKNPPISEVAPGLEDTLILMKIGIA"

```

```

gene
CDs
BASE COUNT      283 a 254 c 283 g 266 t
ORIGIN

```

```

Query Match      23.4%; Score 28.8; DB 8; Length 1086;
Best Local Similarity 58.0%; Pred. No. 24;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 33 ctgatacctaattgagcgaatgagagctgacatgactgtagtcaatgtaagt 92
      || || || || || || || || || || || || || || || || || || || || || ||
Db 735 CTCATCATATATMAAGAACATGCGCAGGGTTCCTTTGTGMAAACCATCATTAATGTC 676

QY 93 gatactatcttcaattatccacaca 120
      || || || || || || || || || || || || || || || || || || || || || ||
Db 675 AATTGTAACTTAAGAAAGCCAGCCCA 648

```

```

RESULT 41
AC018023/c
LOCUS AC018023 14039 bp DNA HTG 09-DEC-1999
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
          pieces.
ACCESSION AC018023
VERSION AC018023.1 GI:6553167
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
          Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
          Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
          Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 14039)
AUTHORS Adams,M. and Venter,J.C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
          Rockville, MD, USA
COMMENT This sequence was identified as CDM:10213067 by the submitter.
          For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..14039
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"

```

```

BASE COUNT      3993 a 3093 c 3021 g 3932 t
ORIGIN
Query Match      23.4%; Score 28.8; DB 2; Length 14039;
Best Local Similarity 58.0%; Pred. No. 28;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

```

```

QY 34 tgaatccctaattgagcgaatgagagctgacatgactgtagtcaatgtaagt 93
      || || || || || || || || || || || || || || || || || || || || || ||
Db 12793 TGGGGAACTGCTGTACMAAGCCAGGCGCTCACTTGTATTAGTCATCAATGAGCAG 12734

QY 94 ataactatcttcaattatccacaca 121
      || || || || || || || || || || || || || || || || || || || || || ||
Db 12733 ACAAAATAGCTGCTTGTGATCCGATTCGA 12706

```

```

RESULT 42
AE001152
LOCUS AE001152 17078 bp DNA BCT 15-DEC-1997
DEFINITION Borrelia burgdorferi (section 38 of 70) of the complete genome.
ACCESSION AE001152 AE000783
VERSION AE001152.1 GI:2688387
KEYWORDS Lyme disease spirochete.
SOURCE Borrelia burgdorferi
ORGANISM Bacteria; Spirochaetales; Spirochaetaceae; Borrelia; Borrelia
          burgdorferi group.
REFERENCE 1 (bases 1 to 17078)
AUTHORS Fraser,C.M., Casjens,S., Huang,W.M., Sutton,G.G., Clayton,R.,
          Lathigra,R., White,O., Ketchum,K.A., Dodson,R., Hickey,E.K.,
          Gwinn,M., Dougherty,B., Tomb,J.F., Fleischmann,R.D., Richardson,D.,

```

TITLE
 Peterson, J., Kerlavage, A.R., Quackenbush, J., Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M.D., Gocayne, J., Venter, J.C. et al.
 Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*
 Nature 390 (6660), 580-586 (1997)

JOURNAL
 MEDLINE
 98065943
 2 (bases 1 to 17078)
 Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.A., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Gimp, M., Dougherty, B., Tomb, J.-F., Fleischmann, R.D., Richardson, D., Peterson, J., Kerlavage, A.R., Quackenbush, J., Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M.D., Gocayne, J.D., Weidman, J., Utterback, T., Watthey, L., McDonald, L., Artiach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B., Smith, H.O. and Venter, J.C.
 Direct Submission
 Submitted (12-DEC-1997) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
 Location/Qualifiers
 1. 17078
 /organism="Borrelia burgdorferi"
 /db_xref="taxon:139"
 361..1725
 /gene="BB0473"
 361..1725
 /gene="BB0473"
 /note="similar to SP:P54181 PID:1256651 GB:AL009126 percent identity: 23.93; identified by sequence similarity: putative"
 /codon_start=1
 /transl_table=1
 /product="conserved hypothetical integral membrane protein"
 /protein_id="AAC66867.1"
 /db_xref="GI:2688416"
 /translation="MSTDKSKTREILNGNIXKVFILSPVITNITQAFEDLDMVGRKAMPALSLAGPVNFFITAIAMGATGSLSKCIGESNRSFSTYAGQILVNFVLSLVETICAFETDLDDLGVGELGKESRYFYVITFPIIMPFSISTYILNAGETILSMITVIFANVFIIDPILIFSPNMGITGAAMATSKLLTVFVFLYRLNAGKILHMDVIDIRSIKEIVNLGLPSTFGQIMVLSFIFNYIVIESPFLAAYGLNTIISLFLPAMGIGGIIISVGNLGAKKVNEVEYLKFKFISALIIILISIVENKQFIRLFTNDELEVNANNYLLITGCGVGLQOVFPGGLSGSRRTIAMIVIRIRMLIRLPVVFIFQYIMENSLGIAFIISNYAILIILIAFTCRIRWAFILIKKK"
 1764..1865
 /gene="BB0474"
 1764..1865
 /gene="BB0474"
 /note="hypothetical protein: identified by Glimmer; putative"
 /codon_start=1
 /transl_table=1
 /product="B. burgdorferi predicted coding region BB0474"
 /protein_id="AAC66869.1"
 /db_xref="GI:2688418"
 /translation="WAYFLINLMFLAKIDFRFFVNLFSVYCLQ"
 complement(1923..2315)
 /gene="BB0475"
 complement(11923..2315)
 /gene="BB0475"
 /note="hypothetical protein: identified by Glimmer; putative"
 /codon_start=1
 /transl_table=1
 /product="B. burgdorferi predicted coding region BB0475"
 /protein_id="AAC66868.1"
 /db_xref="GI:2688417"
 /translation="YMAQKFRKLPLIVISGLIIVFMSCKMSTSTIKENAKETTYLISTIKINOKVEIVNKSCKNLIITLKNSTEDINANSIAIFEGSKTGLIREKLNGLFTFHLKTKINTKRKTLTYIFERK"
 2785..3990
 /gene="BB0476"
 /gene

CDS
 2785..3990
 /gene="BB0476"
 /note="similar to GB:I23125 PID:349407 SP:P50062 PID:I685369 GB:AE000783 percent identity: 100.00; identified by sequence similarity: putative"
 /codon_start=1
 /transl_table=1
 /product="translation elongation factor TU (tuf)"
 /protein_id="AAC66866.1"
 /db_xref="GI:2688415"
 /translation="MKPRIVMAKEVQRTKPHMNVGTIGVHGKTTLPAISYCSKINRAKAKTEDIDNAPEEAKGRTITNARIETANRHRARHDCGADYIKMNTGAQMDAAILVADSGAEPTQKEHLIAQWGIKTIIVELNKDLPCLVEIYEVLAEVKEKFSADPIITIGSAFGAMSNDEPSTCVKELLESMDNFDLPERIDIKPFLAIVEDFESISGRVATGRIGRIIGVQGEVIVGIKETRTKTVGVGEPKILQGOAGNVGLLTVGVGKDIERGQVLSAGVITTPPKFKRASICYLCTKEGGRKPFEPGVRPOFFRTTDTGVVLAEGKEVMWPGNVDIYELLSIAMKNVEFAVREGRTVASGRILEIE"
 4041..4352
 /gene="BB0477"
 4041..4352
 /gene="BB0477"
 /note="similar to GP:1685370 percent identity: 99.03; identified by sequence similarity: putative"
 /codon_start=1
 /transl_table=1
 /product="ribosomal protein S10 (rpsu)"
 /protein_id="AAC66865.1"
 /db_xref="GI:2688414"
 /translation="MIADKIRVRLFSRDYKILDOSASTYKAKAKOIKPIPIPTKIKKTYLRSPHVNKSRQEPFMRKRLIDILEPTSLAMDSLMKLELPAGVEVDIKQ"
 4388..5008
 /gene="BB0478"
 4388..5008
 /gene="BB0478"
 /note="similar to GP:1685371 percent identity: 99.03; identified by sequence similarity: putative"
 /codon_start=1
 /transl_table=1
 /product="ribosomal protein L3 (rplc)"
 /protein_id="AAC66864.1"
 /db_xref="GI:2688413"
 /translation="MLGILGKVKQTOIFQKNGIYVYVYIEOPNITTKKYVDRGYSALIAGVLDKSSKSPKIPKIGYKSLKNDIEPKRYIELGLDSTDGDIKVDVFSVKYVDTGTTRKGFGOGAMKRNHNSGSPSHSGKFRHHLGGTGATTPATPFGTKWAGRMGNQOTIQLLEVLIIDEEKRALVKGAVPGAKGSFVVKKSKK"
 5018..5647
 /gene="BB0479"
 5018..5647
 /gene="BB0479"
 /note="similar to GP:1685372 percent identity: 99.04; identified by sequence similarity: putative"
 /codon_start=1
 /transl_table=1
 /product="ribosomal protein L4 (rpld)"
 /protein_id="AAC66863.1"
 /db_xref="GI:2688412"
 /translation="MERKVFESKDKKEIGTINLDDRVNIEISHSITYNAIKNELSNLRVGTSTKTRSEVSGSKKPKRGKGTGRARVGTGRNVPVIGGIALGPKPDYSYLPKRVKTLAPKSVSLIRADENSFRKVENNVESGKDKDALIITKNASPNGCVVILLGNDQMIKRAKGNIRDKILSFDKLRVVDLTFYAKNLIALSAVKNLNEFIK"
 5653..5964
 /gene="BB0480"
 5653..5964
 /gene="BB0480"
 /note="similar to GP:1685373 percent identity: 100.00; identified by sequence similarity: putative"
 /codon_start=1
 /transl_table=1
 /product="ribosomal protein L23 (rplw)"
 /protein_id="AAC66862.1"
 /db_xref="GI:2688411"
 /gene

Bandon, R.C., Rogers, Y.H., Elizey, R.G., Champe, M., Pfeiffer, B.D.,
 Man, K.H., Doyle, C., Baxter, E.G., Helt, G., Nelson, C.R., Gabor
 Miklos, G.L., Abill, J.F., Agbayani, A., An, H.J.,
 Andrews-Frankkoch, C., Baldwin, D., Bailey, R.M., Basu, A.,
 Baxendale, J., Bayraktaroglu, L., Beasley, E.M., Beeson, K.Y.,
 Benos, P.V., Berman, B.P., Bhandari, D., Bolshakov, S., Borikova, D.,
 Botchan, M.R., Bouck, J., Broksrein, P., Brotlier, P., Burris, K.C.,
 Busam, D.A., Butler, S., Cadieu, E., Center, A., Chandra, I.,
 Cherry, J.M., Cawley, S., Dahle, C., Davenport, L.B., Davies, P.,
 de Pablos, B., Delcher, A., Deng, Z., Mays, A.D., Dew, I., Dietz, S.M.,
 Dodson, K., Doup, L.E., Downes, M., Dugan-Rocha, S., Dunkov, B.C.,
 Dunn, P., Durbin, K.J., Evanselista, C.C., Ferraz, C., Ferltera, S.,
 Fleischmann, W., Foster, C., Gabriellian, A.E., Garg, N.S.,
 Gelbart, W.M., Glasser, K., Glodek, A., Gong, F., Gorrell, J.H., Gu, Z.,
 Guan, P., Harris, M., Harris, N.L., Harvey, D., Helman, T.J.,
 Hernandez, J.F.R., Houck, J., Hostin, D., Houston, K.A., Howland, T.J.,
 Wei, M.H., Idegawa, C., Jalali, M., Kalush, F., Karpen, G.H., Ke, Z.,
 Kennison, J.A., Ketchum, K.A., Kimmel, B.E., Kodira, C.D., Kraft, C.,
 Kravtsov, S., Kulp, D., Lai, Z., Lascko, P., Lei, Y., Levitsky, A.A.,
 Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T.C.,
 McLeod, M.P., McPherson, D., Mekulov, G., Milshina, N.V., Mobarry, C.,
 Morris, J., Mosher, A., Mount, S.M., Moy, M., Murphy, B., Murphy, L.,
 Muzny, D.M., Nelson, D.L., Nelson, D.R., Nelson, K.A., Nixon, K.,
 Nusskern, D.R., Pacleby, J.M., Palazolo, M., Piltman, G.S., Pan, S.,
 Pollard, J.D., Puri, V., Reese, M.G., Reinert, K., Remington, R.,
 Saunders, R.D., Scheeler, F., Shen, H., Shue, B.C., Siden-Kiamos, I.,
 Simpson, M., Skupski, M.P., Smith, T., Spler, E., Spradling, A.C.,
 Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R.,
 Venter, E., Wang, A.H., Wang, X., Wang, Z.Y., Wasserman, D.A.,
 Weinstein, G.M., Weissenbach, J., Williams, S.M., Woodage, T.,
 Worley, K.C., Wu, D., Yang, S., Yao, Q.A., Ye, J., Yen, R.F.,
 Zaveri, J.S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X.H.,
 Zhong, F.N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H.O.,
 Gibbs, R.A., Myers, E.W., Rubin, G.M. and Venter, J.C.
 The genome sequence of *Drosophila melanogaster*
 Science 287 (5461), 2185-2195 (2000)
 20196006
 2 (bases 1 to 49759)
 Adams, M.D., Celniker, S.E., Gibbs, R.A., Rubin, G.M. and Venter, C.J.
 Direct Submission
 Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,

URES	Location/Qualifiers
source	1..49750
	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"
	/chromosome="2L"
	/map="38C1-38C2"
	<8107..>8511
mRNA	/gene="CG17570"
	/product="Cp34146"
	/db_xref="FLYBASE:FBan0017570"
	/db_xref="FLYBASE:FBgn0032948"

/gene="CG17570"
 /map="39E1..39E1"
 /db_xref="FLYBASE:FBanno017570"
 /db_xref="FLYBASE:FBgn0032948"
 /evidence=not_experimental
 8107..8511
 /gene="CG17570"
 /note="CG17570 gene product"
 /codon_start=1
 /db_xref="FLYBASE:FBanno017570"
 /db_xref="FLYBASE:FBgn0032948"
 /evidence=not_experimental
 /protein_id="AAFA5446.1"
 /db_xref="GI:7289775"
 /translation="MKNIVALKLSKSYPLMGCHDEGNIITLADLDLALQOQISANFTIY
 AISKRTPISEKRIIVLTAKRMLHIQMLRRFRILDKOIRRLIGHKDAKRVLEHGFQOEE
 LFEIIVHEHYMVAVRGLSLDRKRNLLAGOPATRP"
 <15476..16286

gene

CDs

BASE COUNT 14487 a 10540 c 10241 g 14491 t

ORIGIN

Query Match 23.4%; Score 28.8; DB 3; Length 49759;
Best Local Similarity 58.0%; Pred. No. 30;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 34 tgatgactaattgagcgaatgcagagctgacgtgtaagttcacaatgtaagtg 93
Db 12852 TGGGGAACGTGCTACAAAGCGAGGCTGTTAGTTAGTCATCAAGTGAGCAG 12793

Qy 94 ataactatctcaattatccacaaca 121
Db 12792 ACAATTAACCTGCGTTGATCCGATTGCA 12765

RESULT 44
AC004735/c
LOCUS
DEFINITION AC004735 74534 bp DNA INV 31-OCT-2000
Drosophila melanogaster. chromosome 2L, region 38A7-38C9, P1 clone
DS04178, complete sequence.
AC004735 AC004119 AC000718 AC000720 AC000721 L81497
AC001816 L81498 AC000722 AC000723 AC001208 L81496 AC000724
AC000725 AC000726 AC000727
AC004735.1 GI:3892961
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

1 (bases 1 to 74534)
Celiker,S.E., Agbayan,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Butenoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
Moshirefi,A.R., Moshirefi,M., Nixon,K., Paclele,J.M., Park,S.,
Pfeiffer,B., Poon,L., Punch,E., Sequeira,A., Sethi,H., Shit,E.,
Svirskas,R.R., Twomey,B., Wan,K.H., Weinburg,T., Zhang,R.,
Zieran,L.L. and Rubin,G.M.
Sequencing of Drosophila chromosome 2L, region 38C
Unpublished
2 (bases 1 to 74534)
Celiker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
Agbayan,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshirefi,A.R.,
Moshirefi,M., Nixon,K., Paclele,J.M., Park,S., Pfeiffer,B., Punch,E.,

Shit,E., Twomey,B., Wan,K.H., Whitejaw,K.R., Yee,A., Zhang,R.,
Zieran,L.L. and Kimmel,B.E.
Direct Submission
Submitted (22-MAY-1998) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Nov 19, 1998 this sequence version replaced gi:3766072
gi:2829852.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (<http://www.fruitfly.org/sequence/>) or send email
to bdg@fruitfly.berkeley.edu.
P1 library locations: 44-50.
FEATURES
source
1. 74534
/organism="Drosophila melanogaster"
/strain="y2; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2L"
/map="38A7-38C9"
/clone="P1_DS04178 (D78)"
/clone_lib="P1 library, partial Sau3a in pAd10acB1T"
BASE COUNT 22626 a 15319 c 15152 g 21437 t

ORIGIN

Query Match 23.4%; Score 28.8; DB 3; Length 74534;
Best Local Similarity 58.0%; Pred. No. 30;
Matches 51; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

Qy 34 tgatgactaattgagcgaatgcagagctgacgtgtaagttcacaatgtaagtg 93
Db 48537 TGGGGAACGTGCTACAAAGCGAGGCTGTTAGTTAGTCATCAAGTGAGCAG 48478

Qy 94 ataactatctcaattatccacaaca 121
Db 48477 ACAATTAACCTGCGTTGATCCGATTGCA 48450

RESULT 45
AL136093
LOCUS
DEFINITION AL136093 85195 bp DNA PRI 30-SEP-2000
Human DNA sequence from clone RP3-472A9 on chromosome 6q13-14.3,
complete sequence.
AL136093
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

1 (bases 1 to 85195)
Matthews,L.
Direct Submission
Submitted (30-SEP-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
requests: clonerequests@sanger.ac.uk
On Oct 1, 2000 this sequence version replaced gi:9931221.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unSURE'
feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: EMBL; EMBL; SW, SWISSPROT; Tr, TREMBL; wp, WORMPEP; Information

On the WormER database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormerp This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre chromosome 6 Mapping group. Further information can be found at

RP3-772A9 is from the library RP3-3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/>
VECTOR: pCYPAC2

IMPORTANT: this sequence is not the entire insert of clone RP3-472A9. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true right end of clone RP3-472A9 is at 85195 in this sequence. The true right end of clone RP1-134A13 is at 100 in this sequence.

FEATURES

Source

1.85195

```
/organism="Homo sapiens"
```

```
/ab_xref="LAXon:9606"
```

$$\gamma_{\text{CH2OH}} = 6.3^\circ$$

```

/imaP- q10 14.0
/c1one="RP3-472A9"

```

```
/clone_lib="RPCI-3"
```

a 14629 c 16180 g

BASE COUNT	24927	a	14629	c	16180	g	29459	t
ORIGIN								

ORIGIN

Query Match

23.4%; Score 28.8; DB 9; Length 85195;

Best Local Similarity 53.68; Pred. No. 30; Watches 60; Concentration 0; Wastages

Matches	60;	Conservative	0;	Mismatches	52;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

9 gtaaatgtgagatgccttaacaacccgtatgacctaatggcgcaatgcyagggctgcactg 68

Db 61218 GCAATTTTAACTTACTTTTCCAGCTTGGATTCTTATCCGGTAACCTTCTAAGTAAGAAATTG 61277

QY 69 actggttaagttcacatgttaagtgatacttatcttcaattatccacaca 120

Db 61278 TTTTCATGATGTCATATGTTCTGAATATATAATTTCATCATTCACACATA 61329

Search completed: March 19, 2002, 10:00:06
Job time: 1710 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2002, 09:34:11 ; Search time 100.26 Seconds

(without alignments)
1051.775 Million cell updates/sec

Title: US-09-684-016-48411

Perfect score: 123

Sequence: 1 caggtatgttaattgtgaga.....tcaattatccacacacaca 123

Scoring table: IDENTITY_NUC

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: N.Geneseq_1101.*
2: /SIDS2/gcgcdata/geneseq/geneseqn/NA1980.DAT.*
3: /SIDS2/gcgcdata/geneseq/geneseqn/NA1981.DAT.*
4: /SIDS2/gcgcdata/geneseq/geneseqn/NA1982.DAT.*
5: /SIDS2/gcgcdata/geneseq/geneseqn/NA1983.DAT.*
6: /SIDS2/gcgcdata/geneseq/geneseqn/NA1984.DAT.*
7: /SIDS2/gcgcdata/geneseq/geneseqn/NA1985.DAT.*
8: /SIDS2/gcgcdata/geneseq/geneseqn/NA1986.DAT.*
9: /SIDS2/gcgcdata/geneseq/geneseqn/NA1987.DAT.*
10: /SIDS2/gcgcdata/geneseq/geneseqn/NA1988.DAT.*
11: /SIDS2/gcgcdata/geneseq/geneseqn/NA1989.DAT.*
12: /SIDS2/gcgcdata/geneseq/geneseqn/NA1990.DAT.*
13: /SIDS2/gcgcdata/geneseq/geneseqn/NA1991.DAT.*
14: /SIDS2/gcgcdata/geneseq/geneseqn/NA1992.DAT.*
15: /SIDS2/gcgcdata/geneseq/geneseqn/NA1993.DAT.*
16: /SIDS2/gcgcdata/geneseq/geneseqn/NA1994.DAT.*
17: /SIDS2/gcgcdata/geneseq/geneseqn/NA1995.DAT.*
18: /SIDS2/gcgcdata/geneseq/geneseqn/NA1996.DAT.*
19: /SIDS2/gcgcdata/geneseq/geneseqn/NA1997.DAT.*
20: /SIDS2/gcgcdata/geneseq/geneseqn/NA1998.DAT.*
21: /SIDS2/gcgcdata/geneseq/geneseqn/NA2000.DAT.*
22: /SIDS2/gcgcdata/geneseq/geneseqn/NA2001.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51.8	42.1	1422	21	AAc44809
2	49.6	40.3	606	21	AAc42911
3	49.6	40.3	913	21	AAc46260
4	49.6	40.3	1204	21	AAc33604
5	49.6	40.3	1272	21	AAc33124
6	47	38.2	1104	21	AAc46355
7	47	38.2	1145	21	AAc45592
8	47	38.2	1148	21	AAc32774
9	34.4	28.0	1082138	21	AAc22305
10	28.8	23.4	910715	20	AAx20248
11	28.4	23.1	582	21	AAc49264

C	12	26.4	21.5	1474	21	AAc45107	Arabidopsis thalia
C	13	26.4	21.5	2347	21	AAz46167	CDNA sequence enco
C	14	26	21.1	2620	22	AAH15633	Human CDNA sequenc
C	15	25.6	20.8	634	20	AAH20838	Poly nucleotide seq
C	16	25.6	20.8	7147	19	AAV52156	Streptococcus pneu
C	17	25.6	20.8	19932	17	AAV46159	Cagi locus. Helic
C	18	25.4	20.7	591	21	AAZ88746	A. dichotoma kabut
C	19	25.4	20.7	2333	22	AAH17997	Human CDNA sequenc
C	20	25.4	20.7	3300	21	AAH92497	Haemophilus Influe
C	21	25.2	20.5	649	22	AAH83942	Callitrix jaccus
C	22	25.2	20.5	3684	22	AAI60262	Human polynucleoti
C	23	25.2	20.5	12412	17	AAI33537	BCG deletion regio
C	24	25	20.3	933	22	AAH32402	Human olfactory re
C	25	25	20.3	2284	20	AAH04441	S. frugiperda immu
C	26	25	20.3	49998	20	AAH23518	Human kidney amino
C	27	24.8	20.2	528	22	AAH13128	Human CDNA clone (
C	28	24.8	20.2	1448	21	AAc39501	Arabidopsis thalia
C	29	24.8	20.2	1782	22	AAH23414	Human SEC5 DNA seq
C	30	24.8	20.2	1970	22	AAH16592	Human CDNA sequenc
C	31	24.8	20.2	2064	16	AAH80535	Paralichthys oliva
C	32	24.8	20.2	2064	16	AAH80533	Paralichthys oliva
C	33	24.8	20.2	2148	14	AAQ46026	Transglutaminase (
C	34	24.8	20.2	2148	14	AAQ54878	Transglutaminase (
C	35	24.8	20.2	2148	16	AAQ80536	Paralichthys oliva
C	36	24.8	20.2	2148	16	AAQ80534	Paralichthys oliva
C	37	24.6	20.0	452	21	AAc59911	Human secreted pro
C	38	24.6	20.0	625	22	AAH26556	DNA encoding human
C	39	24.6	20.0	1864	8	AAH71405	Sequence of ANS-1
C	40	24.6	20.0	2946	22	AAH62796	Shrimp white spot
C	41	24.6	20.0	3787	22	AAH81775	Human differential
C	42	24.6	20.0	68940	20	AAH57351	Human chromosome 6
C	43	24.6	20.0	109519	22	AAH08693	Micromonospora DNA
C	44	24.4	19.8	294	21	AAc75339	Human ORF ORF894
C	45	24.4	19.8	300	21	AAH01147	Human colon cancer

ALIGNMENTS

RESULT	1
ID	AAc44809 standard; DNA; 1422 BP.
XX	
AC	AAc44809;
XX	
DT	18-OCT-2000 (first entry)
XX	
DE	Zea mays DNA fragment SEQ ID NO: 44214.
XX	
KW	Hybridisation assay; genetic mapping; gene expression control;
KW	Protein identification; signal transduction pathway; metabolic;
KW	pathway; promoter; termination sequence; corn; ss.
XX	
OS	Zea mays subsp. mays.
XX	
PN	EP1033405-A2.
XX	
PD	06-SEP-2000.
XX	
PF	25-FEB-2000; 2000EP-0301439.
XX	
PR	25-FEB-1999; 99US-0121825.
PR	05-MAR-1999; 99US-0123180.
PR	09-MAR-1999; 99US-0123548.
PR	23-MAR-1999; 99US-0125788.
PR	25-MAR-1999; 99US-0126264.
PR	29-MAR-1999; 99US-0126785.
PR	01-APR-1999; 99US-0127462.
PR	06-APR-1999; 99US-0128234.
PR	08-APR-1999; 99US-0128714.
PR	16-APR-1999; 99US-0129845.
PR	19-APR-1999; 99US-0130077.
PR	21-APR-1999; 99US-0130449.

```
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
```


PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 42.1%; Score 51.8; DB 21; Length 1422;
Best Local Similarity 83.1%; Pred. No. 6.2e-09;
Matches 59; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 4 gttatgtaattggaatgacctgataaccctgtgacctatgacgcaatgcgaaggctg 63
Db 783 gttatgcaagtgtgagatgacgtacacccagacacctcatgtgcagtgtgaagctg 842

QY 64 cactgactggt 74
Db 843 caagactggt 853

RESULT 2

ID AAC42911 standard: DNA; 606 BP.

XX AAC42911;

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana DNA fragment SEQ ID NO: 37322.

KW Hybridisation assay; genetic mapping; gene expression control;

KW protein identification; signal transduction pathway;

XX metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.

Query Match	Best Local Similarity	Score	DB	Length
Matches 58; Conservative 0; Mismatches 14; Indels 0; Caps 0;	80.6%;	49.6;	21;	606;
OY 4 gtagtgaatagtgtgagatgccttaacacccgagatgaactaatgagcagatcgagggtg 63				
Db 342 gtaccgaatgttgaatgccttaacatccagacagatcatgtgtcagttgaaagctg 401				
OY 64 cactgactgcta 75				
Db 402 caaagactgcta 413				
RESULT 3				
AAAC46260				
ID AAC46260 standard; DNA; 913 BP.				
AC AAC46260;				
XX 18-OCT-2000 (first entry)				
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 49491.				
XX				
KW Hybridisation assay; genetic mapping; gene expression control;				
KM protein identification; signal transduction pathway;				
RW metabolic pathway; promoter; termination sequence; ss.				
XX				
OS Arabidopsis thaliana.				
XX				
PN EPI033405-A2.				
XX				
PD 06-SEP-2000.				
XX				
PF 25-FEB-2000; 2000EP-0301439.				
XX				
PR 25-FEB-1999; 99US-0121825.				
PR 05-MAR-1999; 99US-0123180.				
PR 09-MAR-1999; 99US-0123548.				
PR 23-MAR-1999; 99US-0125788.				
PR 25-MAR-1999; 99US-0126254.				
PR 29-MAR-1999; 99US-0126785.				
PR 01-APR-1999; 99US-0127462.				
PR 06-APR-1999; 99US-0128234.				
PR 08-APR-1999; 99US-0128714.				
PR 16-APR-1999; 99US-0129845.				
PR 19-APR-1999; 99US-0130077.				
PR 21-APR-1999; 99US-0130549.				
PR 23-APR-1999; 99US-0130891.				
PR 28-APR-1999; 99US-0131449.				
PR 30-APR-1999; 99US-0132048.				
PR 30-APR-1999; 99US-0132407.				
PR 04-MAY-1999; 99US-0132484.				

```
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 07-MAY-1999; 9905-0132487.
PR 11-MAY-1999; 9905-0132863.
PR 14-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 18-MAY-1999; 9905-0134370.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 02-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.

PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148319.
PR 11-AUG-1999; 9905-0148341.
PR 12-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155386.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158332.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
```

```
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160984.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 40.3%; Score 49.6; DB 21; Length 913;
Best Local Similarity 80.6%; Pred. No. 3.3e-08;
Matches 58; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 4 gttatgttaattgtgagtgccttacaacctgtgacctgaatgctgcaatgcgaggtctg 63
Db 344 gttatgttaattgtgagtgccttacaacctgtgacctgaatgctgcaatgcgaggtctg 403
Qy 64 cactgactgtgta 75
Db 404 caaagactgtgta 415

RESULT 4
AAC33604
ID AAC33604 standard; DNA; 1204 BP.
XX AAC33604;
AC 17-OCT-2000 (first entry)
XX
DT
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 3658.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
XX Arabidopsis thaliana.
OS
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131448.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
```

```
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
```


PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151067.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158363.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160776.
PR 21-OCT-1999; 99US-0160777.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.

```
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161355.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161923.
PR 29-OCT-1999; 99US-0162142.

Query Match 40.3%; Score 49.6; DB 21; Length 1272;
Best Local Similarity 80.6%; Pred. No. 3.7e-08;
Matches 58; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Oy 4 gatatgtaattgagatgccttaacacctatgacctaatggcgcaatgcgagggctg 63
Db 701 gtactgcaaatgtgaattgccttaacatccagacgactcatcgtgctgagtgagggctg 760
Oy 64 cactgactgta 75
Db 761 caaagactgta 772

RESULT 6
AAC46355
ID AAC46355 standard; DNA; 1104 BP.
XX
AC AAC46355;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 49837.
XX
KM Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0131491.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
```

```
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143342.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
```

PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151086.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-015659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159285.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.

PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 38.2%; Score 47; DB 21; Length 1104;
Best Local Similarity No. 78; 98; Pred. No. 3e-07;
Matches 56; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 4 gttctgtaaatgtgagatgcttaacacctgatacctaattgycgcgaatgcgagggctg 63
Db 837 gtctcgcaagtgtgagatgcttaacacctgatacctaattgctgattgtgcaatgcgagggctg 896

Qy 64 cactgactgtgt 74
Db 897 ttcttgatgtgt 907

RESULT 7
AAC45592
ID AAC45592 standard; DNA; 1145 BP.
XX
AC AAC45592;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 47055.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999;	99US-0135353.
PR 24-MAY-1999;	99US-0135629.
PR 25-MAY-1999;	99US-0136021.
PR 27-MAY-1999;	99US-0136392.
PR 28-MAY-1999;	99US-0136782.
PR 01-JUN-1999;	99US-0137222.
PR 03-JUN-1999;	99US-0137528.
PR 04-JUN-1999;	99US-0137502.
PR 07-JUN-1999;	99US-0137724.
PR 08-JUN-1999;	99US-0138094.
PR 10-JUN-1999;	99US-0138540.
PR 10-JUN-1999;	99US-0138847.
PR 14-JUN-1999;	99US-0139119.
PR 16-JUN-1999;	99US-0139452.
PR 16-JUN-1999;	99US-0139453.
PR 17-JUN-1999;	99US-0139492.
PR 18-JUN-1999;	99US-0139454.
PR 18-JUN-1999;	99US-0139455.
PR 18-JUN-1999;	99US-0139456.
PR 18-JUN-1999;	99US-0139457.
PR 18-JUN-1999;	99US-0139458.
PR 18-JUN-1999;	99US-0139459.
PR 18-JUN-1999;	99US-0139460.
PR 18-JUN-1999;	99US-0139461.
PR 18-JUN-1999;	99US-0139462.
PR 18-JUN-1999;	99US-0139463.
PR 18-JUN-1999;	99US-0139750.
PR 18-JUN-1999;	99US-0139763.
PR 21-JUN-1999;	99US-0139817.
PR 22-JUN-1999;	99US-0139899.
PR 23-JUN-1999;	99US-0140353.
PR 23-JUN-1999;	99US-0140354.
PR 24-JUN-1999;	99US-0140695.
PR 28-JUN-1999;	99US-0140823.
PR 29-JUN-1999;	99US-0140991.
PR 30-JUN-1999;	99US-0141287.
PR 01-JUL-1999;	99US-0141842.
PR 01-JUL-1999;	99US-0142154.
PR 02-JUL-1999;	99US-0142055.
PR 06-JUL-1999;	99US-0142390.
PR 08-JUL-1999;	99US-0142803.
PR 09-JUL-1999;	99US-0142920.
PR 12-JUL-1999;	99US-0142977.
PR 13-JUL-1999;	99US-0143542.
PR 14-JUL-1999;	99US-0143624.
PR 15-JUL-1999;	99US-0144005.
PR 16-JUL-1999;	99US-0144085.
PR 16-JUL-1999;	99US-0144086.
PR 19-JUL-1999;	99US-0144325.
PR 19-JUL-1999;	99US-0144331.
PR 19-JUL-1999;	99US-0144332.
PR 19-JUL-1999;	99US-0144333.
PR 19-JUL-1999;	99US-0144334.
PR 20-JUL-1999;	99US-0144335.
PR 20-JUL-1999;	99US-0144352.
PR 20-JUL-1999;	99US-0144632.
PR 20-JUL-1999;	99US-0144884.
PR 21-JUL-1999;	99US-0144814.
PR 21-JUL-1999;	99US-0145086.
PR 21-JUL-1999;	99US-0145088.
PR 22-JUL-1999;	99US-0145085.
PR 22-JUL-1999;	99US-0145087.
PR 22-JUL-1999;	99US-0145089.
PR 22-JUL-1999;	99US-0145192.
PR 23-JUL-1999;	99US-0145145.
PR 23-JUL-1999;	99US-0145218.
PR 23-JUL-1999;	99US-0145224.
PR 26-JUL-1999;	99US-0145276.
PR 27-JUL-1999;	99US-0145913.
PR 27-JUL-1999;	99US-0145918.
PR 27-JUL-1999;	99US-0145919.
PR 28-JUL-1999;	99US-0145951.
PR 02-AUG-1999;	99US-0146386.
PR 02-AUG-1999;	99US-0146388.
PR 02-AUG-1999;	99US-0146389.
PR 03-AUG-1999;	99US-0147038.
PR 04-AUG-1999;	99US-0147204.
PR 04-AUG-1999;	99US-0147302.
PR 05-AUG-1999;	99US-0147192.
PR 05-AUG-1999;	99US-0147260.
PR 06-AUG-1999;	99US-0147303.
PR 06-AUG-1999;	99US-0147416.
PR 09-AUG-1999;	99US-0147493.
PR 09-AUG-1999;	99US-0147935.
PR 10-AUG-1999;	99US-0148171.
PR 11-AUG-1999;	99US-0148319.
PR 12-AUG-1999;	99US-0148341.
PR 13-AUG-1999;	99US-0148565.
PR 13-AUG-1999;	99US-0148684.
PR 16-AUG-1999;	99US-0149368.
PR 17-AUG-1999;	99US-0149175.
PR 18-AUG-1999;	99US-0149426.
PR 20-AUG-1999;	99US-0149722.
PR 20-AUG-1999;	99US-0149723.
PR 20-AUG-1999;	99US-0149929.
PR 23-AUG-1999;	99US-0149902.
PR 23-AUG-1999;	99US-0149930.
PR 25-AUG-1999;	99US-0150566.
PR 26-AUG-1999;	99US-0150884.
PR 27-AUG-1999;	99US-0151065.
PR 27-AUG-1999;	99US-0151066.
PR 27-AUG-1999;	99US-0151080.
PR 30-AUG-1999;	99US-0151303.
PR 31-AUG-1999;	99US-0151438.
PR 01-SEP-1999;	99US-0151930.
PR 07-SEP-1999;	99US-0152363.
PR 10-SEP-1999;	99US-0153070.
PR 13-SEP-1999;	99US-0153758.
PR 15-SEP-1999;	99US-0154018.
PR 16-SEP-1999;	99US-0154039.
PR 20-SEP-1999;	99US-0154779.
PR 22-SEP-1999;	99US-0155139.
PR 23-SEP-1999;	99US-0155486.
PR 24-SEP-1999;	99US-0155659.
PR 28-SEP-1999;	99US-0156458.
PR 29-SEP-1999;	99US-0156596.
PR 04-OCT-1999;	99US-0157117.
PR 05-OCT-1999;	99US-0157753.
PR 06-OCT-1999;	99US-0157865.
PR 07-OCT-1999;	99US-0158029.
PR 08-OCT-1999;	99US-0158232.
PR 12-OCT-1999;	99US-0158369.
PR 13-OCT-1999;	99US-0159293.
PR 13-OCT-1999;	99US-0159294.
PR 13-OCT-1999;	99US-0159295.
PR 14-OCT-1999;	99US-0159329.
PR 14-OCT-1999;	99US-0159330.
PR 14-OCT-1999;	99US-0159331.
PR 14-OCT-1999;	99US-0159637.
PR 14-OCT-1999;	99US-0159638.
PR 18-OCT-1999;	99US-0159584.
PR 21-OCT-1999;	99US-0160741.
PR 21-OCT-1999;	99US-0160767.
PR 21-OCT-1999;	99US-0160768.
PR 21-OCT-1999;	99US-0160770.
PR 21-OCT-1999;	99US-0160814.
PR 21-OCT-1999;	99US-0160815.
PR 22-OCT-1999;	99US-0160980.
PR 22-OCT-1999;	99US-0160981.
PR 22-OCT-1999;	99US-0160989.
PR 25-OCT-1999;	99US-0161404.
PR 25-OCT-1999;	99US-0161405.
PR 25-OCT-1999;	99US-0161406.
PR 26-OCT-1999;	99US-0161359.
PR 26-OCT-1999;	99US-0161360.
PR 26-OCT-1999;	99US-0161361.

PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
Query Match 38.28; Score 47; DB 21; Length 1145;
Best Local Similarity 78.98; Pred. No. 3,1e-07;
Matches 56; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
OY 4 gttctgaatgtgtagaccccttaacacccctgataccttaatgcgcgaatgcgagctg 63
Db 566 gttctgcacagtgtagagatgcgcgtataaccacgatactgtatgctgcgaatgcgagagctg 625
OY 64 cactgactgagt 74
Db 626 ttctcgagtgt 636
RESULT 8
AAC32774
ID AAC32774 standard; DNA; 1148 BP.
XX AAC32774;
AC AAC32774;
XX 17-OCT-2000 (first entry)
DT 17-OCT-2000 (first entry)
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 602.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144088.
PR 16-JUL-1999; 99US-0144088.
PR 16-JUL-1999; 99US-0144088.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144333.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.

PR	04-AUG-1999	99US-01472702
PR	04-AUG-1999	99US-01471792
PR	05-AUG-1999	99US-01471760
PR	05-AUG-1999	99US-01473703
PR	06-AUG-1999	99US-01474716
PR	06-AUG-1999	99US-01474763
PR	09-AUG-1999	99US-01479735
PR	10-AUG-1999	99US-01481711
PR	11-AUG-1999	99US-01483119
PR	12-AUG-1999	99US-01485615
PR	13-AUG-1999	99US-01486864
PR	16-AUG-1999	99US-01493758
PR	17-AUG-1999	99US-01491765
PR	18-AUG-1999	99US-01494726
PR	20-AUG-1999	99US-01497422
PR	20-AUG-1999	99US-01497223
PR	23-AUG-1999	99US-01499302
PR	23-AUG-1999	99US-01499302
PR	25-AUG-1999	99US-01505666
PR	26-AUG-1999	99US-01508844
PR	27-AUG-1999	99US-01510065
PR	27-AUG-1999	99US-01510066
PR	27-AUG-1999	99US-01510080
PR	30-AUG-1999	99US-01510103
PR	31-AUG-1999	99US-01514338
PR	01-SEP-1999	99US-01519130
PR	07-SEP-1999	99US-01552363
PR	10-SEP-1999	99US-01530700
PR	13-SEP-1999	99US-01537358
PR	15-SEP-1999	99US-01540418
PR	16-SEP-1999	99US-01540339
PR	20-SEP-1999	99US-01551779
PR	22-SEP-1999	99US-01551739
PR	22-SEP-1999	99US-01551836
PR	24-SEP-1999	99US-01556458
PR	28-SEP-1999	99US-01564558
PR	28-SEP-1999	99US-01565616
PR	04-OCT-1999	99US-01571117
PR	05-OCT-1999	99US-01577533
PR	06-OCT-1999	99US-01578665
PR	07-OCT-1999	99US-01580829
PR	08-OCT-1999	99US-01583232
PR	12-OCT-1999	99US-01583639
PR	13-OCT-1999	99US-01592933
PR	13-OCT-1999	99US-01592954
PR	14-OCT-1999	99US-01593229
PR	14-OCT-1999	99US-01593229
PR	14-OCT-1999	99US-01593311
PR	14-OCT-1999	99US-01596337
PR	14-OCT-1999	99US-01596337
PR	18-OCT-1999	99US-01595844
PR	21-OCT-1999	99US-01607441
PR	21-OCT-1999	99US-01607657
PR	21-OCT-1999	99US-01607678
PR	21-OCT-1999	99US-01607770
PR	21-OCT-1999	99US-01608114
PR	21-OCT-1999	99US-01608115
PR	22-OCT-1999	99US-01609881
PR	22-OCT-1999	99US-01609881
PR	22-OCT-1999	99US-01609889
PR	25-OCT-1999	99US-01614004
PR	25-OCT-1999	99US-01614005
PR	25-OCT-1999	99US-01614005
PR	26-OCT-1999	99US-01613569
PR	26-OCT-1999	99US-01613561
PR	28-OCT-1999	99US-01613620
PR	28-OCT-1999	99US-01613932
PR	28-OCT-1999	99US-01619933

PR	29-OCT-1999;	99US-0162142.
Query Match	38.2%;	Score 47; DB 21; Length 1148;
Best Local Similarity	78.9%;	Pred. No. 3.1e-07;
Matches	56; Conservative	0; Mismatches 15; Indels 0; Gaps 0;
OY	4 gttatgttaagtgtgagatgccttaaacccgatgaactctaattggcgcgaatgcgaggctg	63
Dd	568 gtctcgcaagaigtgtgagatgccgtataaccacgatcatgttgatgtgcgaatgcgaggatg	627
OY	64 cacttacttgt 74	
Dd	628 ttcttgatcgt 638	

RESULT	9
ID	AAF22305 standard; DNA; 1082138 BP.
XX	
AC	AAF22305;
XX	
DT	20-MAR-2001 (first entry)
XX	
DE	Arabidopsis thaliana chromosome 4 centromere.
XX	
KW	Centromere; microsome; vector; ds.
XX	
OS	Arabidopsis thaliana.
XX	
PN	WO200055325-A2.
XX	
PD	21-SEP-2000.
XX	
PE	17-MAR-2000; 2000WO-US07392.
XX	
PR	18-MAR-1999; 99US-0125219.
PR	01-APR-1999; 99US-0127409.
PR	18-MAY-1999; 99US-0134770.
PR	13-SEP-1999; 99US-0153584.
PR	17-SEP-1999; 99US-0154603.
PA	(UYCH-) UNIV CHICAGO.
XX	
PI	Preuss D, Copenhaver G, Keith K;
XX	
DR	WPI; 2000-587529/55.
XX	
PT	Recombinant DNA construct comprising a plant centromere, useful for
PT	producing stably inherited microsomes which can serve as vectors for
PT	the construction of transgenic plant and animal cells -
XX	
PS	Claim 68; Page 977-1388; 1449pp; English.
XX	
CC	The present invention relates to a recombinant DNA construct of a plant
CC	(Arabidopsis thaliana) centromere. The constructs are useful for
CC	producing stably inherited microsomes which can serve as vectors for
CC	the construction of transgenic plant and animal cells expressing
CC	selected proteins such as hormones, enzymes, interleukins, clotting
CC	factors, cytokines, antibodies, and growth factors.
SQ	Sequence 1082138 BP; 348775 A; 194404 C; 195515 G; 343444 T; 0 other;
XY	
Query Match	28.0%; Score 34.4; DB 21; Length 1082138;
Best Local Similarity	78.8%; Pred. No. 0.11; Indels 0; Gaps
Matches 41; Conservative 0; Mismatches 11;	
OY	2 agctatgtaaatgtgaagtgccttacacaacctgatgactcaatggcgcaat 53 Db 83890 agttactacaatatgtgaattgcgtataaccacagatgactcatgtgccaat 83941
RESULT	10

PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.

PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151348.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 23.1%; Score 28.4; DB 21; Length 582;
Best Local Similarity 76.1%; Pred. No. 1.2; Indels 0; Gaps 0;
Matches 35; Conservative 0; Mismatches 11;

Oy 4 gttatgtaaatgtagatgccttacaacctgtagcctaattgagc 49
||| | ||||| ||||| ||| | ||||| |||||
Db 327 gtactacaatgtgaatgctgatactaccagatgactgagtggag 372

RESULT 12
AAC45107/c
ID AAC45107 standard; DNA; 1474 BP.
XX
AC AAC45107;
XX
XX 18-OCT-2000 (first entry)
DT
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 45325.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX

OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139839.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145066.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145102.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147483.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.

Query Match	21.5%	Score 26.4	DB 21	Length 1474
Best Local Similarity	61.8%	Pred. No. 8.8		
Matches 42	Conservative 0	Mismatches 26	Indels 0	Gaps 0
QY 55	cgagggcgcgcactgactgtgaattccacatgtaagtgatcactatctcaattatc	114		
Db 937	CGATGGCCTCCTGACTGAGGTCTTCTCGCGTGTGTCTGCCACACATTTCTTC	878		
QY 115	cacacac 122			
Db 877	ACCACAC 870			
RESULT 13				
AAZ46167				
ID AAZ46167	standard; cDNA; 2347 BP.			
XX				
AC AAZ46167;				
XX				
DT 16-MAY-2000	(first entry)			
XX				
DE cDNA sequence encoding a human phosphorylation effector PHSP-30.				
XX				
KW Human: phosphorylation effector; PHSP; proliferative disorder;				
KW Immune disorder; neuronal disorder; ss.				
XX				
OS Homo sapiens.				
XX				
Key Location/Qualifiers				
FT CDS 62..1132				

[illegible]

PN EP1074617-A2.
PD 07-FEB-2001.
XX 28-JUL-2000; 2000EP-0116126.
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 13969; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH1633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2620 BP; 788 A; 581 C; 456 G; 795 T; 0 other;

Query Match 21.1%; Score 26; DB 22; Length 2620;
Best Local Similarity 52.8%; Pred. No. 15;
Matches 56; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

OY 5 tattgtaaatgtgagatgacctacacacctgaatgaatgagcgaatgagggctgc 64
DB 1339 TAGAGCAGCAGGAGAAAGCTTTAAACCTGCGCTTATGGAGACGTTACACGAGAGATGCG 1280
OY 65 acgacgtgtaagtctcacatgtaaggtgatactatctcaatt 110
DB 1279 ACAGAGTGAGAGATCCATTTTAAAGCTGATGATGATTAATATT 1234

RESULT 15
AAAX20838
ID AAX20838 standard; DNA: 634 BP.
XX
AC AAX20838;
XX
DT 05-MAY-1999 (first entry)
XX

DE Polynucleotide sequence from the genome of Treponema pallidum.
XX
XX Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW enzyme production; ds.
XX
OS Treponema pallidum.
XX
PN WO9859034-A2.
XX
PD 30-DEC-1998.
XX
PF 23-JUN-1998; 98WO-US13041.
XX
PR 24-JUN-1997; 97US-0050667.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Fraser CM;
XX
DR WPI; 1999-081273/07.
XX
XX
PT New isolated Treponema pallidum nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of T. pallidum infections, particularly syphilis
XX
XX Claim 1; Page 929; 1150pp; English.
XX
XX AAX20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.
XX
SQ Sequence 634 BP; 202 A; 127 C; 148 G; 155 T; 2 other;

Query Match 20.8%; Score 25.6; DB 20; Length 634;
Best Local Similarity 62.5%; Pred. No. 13;
Matches 40; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

OY 49 gaaatgcgaaggctgactgtaagttaacatgtaagtgatgatacttcaat 108
DB 348 gctaaagaaatctgcactctctactcaatttacttggaaacccaactgtcttaaaa 407
OY 109 ttat 112
DB 408 atat 411

RESULT 16
AAV52156/C
ID AAV52156 standard; DNA: 7147 BP.
XX
AC AAV52156;
XX
DT 23-OCT-1998 (first entry)
XX
DE Streptococcus pneumoniae genome fragment SEQ ID NO:23.
XX
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
XX
OS Streptococcus pneumoniae.
XX
PN WO9818931-A2.
XX
PD 07-MAY-1998.
XX
PF 30-OCT-1997; 97WO-US19588.
XX
PR 31-OCT-1996; 96US-0029960.
XX

PA (HUMA-) HUMAN GENOME SCI INC.
XX Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M.
PI Kunsch CA, Rosen CA.
XX WPI: 1998-272225/24.
DR
XX
XX
XX Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
PS
XX
XX
XX Claim 1: Page 289-293; 1409pp; English.
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
CC recorded on it, or a representative fragment or a sequence at least 958
CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
CC Streptococcus pneumoniae. The present invention also describes an
CC isolated nucleic acid molecule encoding a homologue of any of the
CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
CC nucleic acid molecule is produced by a process comprising: (a) screening
CC a genomic DNA library using as a probe a target sequence defined by any
CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC library which contain sequences that hybridise to the target sequence and
CC isolating the nucleic acid molecules from the members; or (b) isolating
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC molecules whose nucleotide sequence is homologous to amplification
CC primers derived from the fragment of the S. pneumoniae genome to prime
CC the amplification and isolating the amplified sequences. The computer
CC readable medium can be used in a computer-based system for identifying
CC fragments of the S. pneumoniae genome of commercial importance, or
CC expression modulating fragments of the S. pneumoniae genome. Products
CC from the present invention can be used in diagnosis kits and assays, and
CC pharmaceutical compositions and vaccines for S. pneumoniae.
SQ Sequence 7147 BP; 2088 A; 1641 C; 1265 G; 2153 T; 0 other;

Query Match 20.8%; Score 25.6; DB 19; Length 7147;
Best Local Similarity 70.8%; Pred. No. 29;
Matches 34; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 28 caacctgatagcctaagcgccatgagcgagcgctgactgctgta 75
DB 219 CAACATATGAGGAACATGATGCGCAAGTGTCTTACGCTGACTGTGA 172

RESULT 17
AAAT46159
ID AAAT46159 standard; DNA; 19932 BP.
XX
AC AAAT46159;
XX
XX 30-JUN-1997 (first entry)
XX
XX CagI locus.
XX
XX CagI; CagA: virulence factor; exporter molecule; homology; pLI gene;
KW Bordetella pertussis; YJR B4; Agrobacterium tumefaciens; invasion factor;
KW Salmonella: type I strain; virulence; diagnosis; H. pylori; infection;
KW vaccine; treatment; duodenal; gastric ulcer; active gastritis;
KW adenocarcinoma; ss.
XX
XX Helicobacter pylori.
XX
XX
XX Key location/Qualifiers
FH CDS 389..1579
FT /tag= a
FT /note= "putative open reading frame; no start codon
FT given"
FT CDS 1794..2168
FT

FT /tag= b
FT /note= "putative open reading frame; no start codon
FT given"
FT 3343..4491
FT /tag= c
FT /note= "putative open reading frame; no start codon
FT given"
FT 4488..5426
FT /tag= d
FT /note= "putative open reading frame; no start codon
FT given"
FT complement (1178..2173)
FT /tag= e
FT /note= "putative open reading frame; no start codon
FT given"
FT complement (2591..3001)
FT /tag= f
FT /note= "putative open reading frame; no start codon
FT given"
FT complement (3020..3259)
FT /tag= g
FT /note= "putative open reading frame; no start codon
FT given"
FT complement (5508..6233)
FT /tag= h
FT /note= "putative open reading frame; no start codon
FT given"
FT complement (6218..6613)
FT /tag= h
FT /note= "putative open reading frame; no start codon
FT given"
FT complement (6562..7074)
FT /tag= i
FT /note= "putative open reading frame; no start codon
FT given"
FT complement (7007..7375)
FT /tag= j
FT /note= "putative open reading frame; no start codon
FT given"
FT complement (7371..7802)
FT /tag= k
FT /note= "putative open reading frame; no start codon
FT given"
FT complement (7975..8454)
FT /tag= l
FT /note= "putative open reading frame; no start codon
FT given"
FT complement (8496..8918)
FT /tag= m
FT /note= "putative open reading frame; no start codon
FT given"
FT complement (8977..9762)
FT /tag= n
FT /note= "putative open reading frame; no start codon
FT given"
FT complement (9784..10575)
FT /tag= o
FT /note= "putative open reading frame; no start codon
FT given"
FT 10557..12758
FT /tag= p
FT /note= "putative open reading frame; no start codon
FT given"
FT complement (12743..13420)
FT /tag= q
FT /note= "putative open reading frame; no start codon
FT given"
FT complement (13374..13742)
FT /tag= r
FT /note= "putative open reading frame; no start codon
FT given"
FT complement (13866..14129)
FT /tag= s
FT

```
FT /note= "putative open reading frame; no start codon
FT given"
FT CDS complement (19296..19832)
FT /*tag= t
FT /note= "putative open reading frame; no start codon
FT given"
XX WO9633274-A1.
XX
XX 24-OCT-1996.
XX
XX
XX 18-APR-1996; 96NO-IB00343.
XX
XX 07-JUN-1995; 95US-0477451.
XX
XX 20-APR-1995; 95US-0425194.
XX
XX (BIOC-) BIOGINE SPA.
XX
XX Covacci A;
XX
XX WPI; 1996-485780/48.
XX
XX P-PSDB; AAM06930-50.
XX
XX Helicobacter pylori CagI polynucleotide and related proteins - used
XX in diagnosis and in vaccines for the treatment of H. pylori
XX infection associated disease
XX
XX Claim 2; Fig 3A-R; 303pp; English.
XX
XX The present sequence is the complete nucleotide sequence of the CagI
XX locus, including the true 5'-terminus of CagI. The sequence was
XX constructed using overlapping clones (one of which also overlaps with
XX the Caga region). The CagI region contains clusters of putative open
XX reading frames (ORFs) with different polarities. The putative ORFs for
XX this region are shown above. It is hypothesised that some of these ORFs
XX may encode exporter molecules with homology to the pil genes of
XX Bordetella pertussis and VIR B4 genes of Agrobacterium tumefaciens and
XX for proteins with motifs shared by the purported invasion factors of
XX Salmonella genus. The absence of the Caga gene in the type I strains is
XX associated with the absence of CagI sequences (which may encode virulence
XX factors restricted to type I strains). The CagI nucleotide sequence, its
XX fragments and encoded proteins are used in the diagnosis of H. pylori
XX (esp. H. pylori type I strain) infection in an individual and in vaccines
XX (claimed) for the treatment of H. pylori infection associated with e.g.
XX duodenal and gastric ulcers, severe forms of active gastritis (esp. type
XX gastritis) and gastric adenocarcinoma.
XX
XX Sequence 19932 BP; 6677 A; 3635 C; 3483 G; 6137 T; 0 other;
XX
XX
XX Query Match 20.8%; Score 25.6; DB 17; Length 19932;
XX Best Local Similarity 51.8%; Pred. No. 42;
XX Matches 58; Conservative 0; Mismatches 54; Indels 0; Gaps 0;
XX
XX 11 aaatgtgagatgccttaacaccgcgagtgactaattgagcgaatgagagtcgactac 70
XX ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX
XX 8085 aaatgcttgaatctgactataacgacccaatttcgcgccttgctgtgccttgag 8144
XX ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX
XX 71 tggtaagttcacatgtaagtgactactactcattcaattatccacacaac 122
XX | ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX
XX 8145 cgttaaatccctcgttaattgtctaataattgaatattgttcccccacagc 8196
XX
XX
XX RESULT 18
XX AA28746
XX ID AA288746 standard; DNA: 591 BP.
XX
XX AA288746;
XX
XX 18-MAY-2000 (first entry)
XX
XX A. dichotoma kabutოსin A DNA.
XX
```

```
KW Kabutოსin A; kabutოსin B; antibacterial; food preservative;
KW medical; ds.
XX
XX Allomyrina dichotoma.
XX
XX Key Location/Qualifiers
XX CDS 82..516
XX /*tag= a
XX /product= "kabutოსin A"
XX
XX JP2000026499-A.
XX
XX 25-JAN-2000.
XX
XX 13-JUL-1998; 98JP-0197190.
XX
XX 13-JUL-1998; 98JP-0197190.
XX
XX (NORO) NORINSUISANSO SANSHI KONCHU.
XX
XX (NODA) ZH NODA SANGYO KAGAKU KENKYUSHO.
XX
XX WPI; 2000-176990/16.
XX
XX P-PSDB; AAY51547.
XX
XX A new peptide, an antibacterial agent, a new peptide gene, a new
XX recombinant DNA and preparation of a new peptide
XX
XX Example 2; Page 9-10; 17pp; Japanese.
XX
XX This invention describes two novel peptides which have antibacterial
XX activity, designated kabutოსin A and B and which have been isolated from
XX Allomyrina dichotoma. The peptides are used as food preservatives
XX and medical antibacterial agents. This sequence encodes the
XX A. dichotoma kabutოსin A protein described in the method of the
XX invention.
XX
XX Sequence 591 BP; 183 A; 123 C; 119 G; 166 T; 0 other;
XX
XX
XX Query Match 20.7%; Score 25.4; DB 21; Length 591;
XX Best Local Similarity 58.7%; Pred. No. 15;
XX Matches 44; Conservative 0; Mismatches 31; Indels 0; Gaps 0;
XX
XX 23 ccttaaaccttgatgactaattgagcgaatgagagtcgactgtaagttca 82
XX ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX
XX 299 ccttaacgacagtgactaattcccgatcccggtcacagtactaccacacatca 358
XX
XX 83 catgtaagtgatgac 97
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX
XX 359 cctctaactgtgaac 373
XX
XX
XX RESULT 19
XX AAH17997/C
XX ID AAH17997 standard; CDNA: 2333 BP.
XX
XX AAH17997;
XX
XX 26-JUN-2001 (first entry)
XX
XX Human cDNA sequence SEQ ID NO:17797.
XX
XX Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
XX Homo sapiens.
XX
XX EP1074617-A2.
XX
XX 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
```

PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR WPI: 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 17797; 2537bp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to
CC AAH95893 represent human amino acid sequences; and AAH13639 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 2333 BP; 699 A; 427 C; 416 G; 791 T; 0 other;
XX
Query Match 20.7%; Score 25.4; DB 22; Length 2333;
Best Local Similarity 56.6%; Pred. No. 24;
Matches 47; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
OY 27 acaacctatgacctaagcgcaatgagagcgctgactgtaagttcaatg 86
DB 1547 ACTGCACAGACCCCTGTACTGGCACAAGTTGATGCTCACTGGAAAAAATGCCAATA 1488
OY 87 taaggtgatactatctcaatt 109
DB 1487 TAAAGTCGAAGTCAGATTCAAAT 1465
XX
RESULT 20
AAA92497
ID AAA92497 standard; DNA; 3300 BP.
XX
AC AAA92497;
XX
DT 17-JAN-2001 (first entry)
XX
DE Haemophilus influenzae adhesin (Hia) gene from NTH1 strain K9.
XX
KW Hia: adhesin; Haemophilus influenzae adhesin; NTH1: infection; vaccine;
KW non-typable Haemophilus influenzae; antiinflammatory; auditory;
KW antibacterial; meningitis; epiglottitis; septicemia; otitis media;
KW diagnosis; immunogenic; antigen; ds.
XX

OS Haemophilus influenzae.
XX
PN WO200055191-A2.
XX
PD 21-SEP-2000.
XX
XX 16-MAR-2000; 2000WO-CA00289.
PF
XX 16-MAR-1999; 99US-0268347.
PR
XX (CONN-) CONNAUGHT LAB LTD.
PA
PI Loosmore SM, Yang Y, Klein MH;
PI WPI: 2000-618897/59.
DR P-PSDB; AAB23858.
DR
XX
PT Novel nucleic acid encoding Hemophilus influenzae adhesin protein, for
PT use as antigens and vaccines and for treating Hemophilus influenzae
PT infection -
XX
PS Claim 1; Fig 22; 275bp; English.
XX
XX The present sequence represents a Haemophilus influenzae adhesin (Hia)
CC gene from the non-typable Haemophilus influenzae (NTH1) strain K9.
CC Hia genes and proteins have antiinflammatory, auditory and antibacterial
CC activities, and can be used in the production of a vaccine. An
CC immunogenic composition comprising an Hia gene, a polypeptide encoded
CC by an Hia gene, or a recombinant Hia polypeptide is useful for inducing
CC protection against disease caused by Haemophilus strains in a
CC susceptible host, preferably a human. An Hia protein is useful as an
CC antigen, in immunogenic preparations including vaccines, as a carrier
CC for other immunogens, and in the generation of diagnostic reagents. Hia
CC is useful for treating diseases caused by the infection of Haemophilus
CC influenzae such as meningitis, epiglottitis, septicemia and otitis
CC media. Recombinant production of Hia favours high recovery of the
CC protein compared to the low recovery of native protein from Haemophilus
CC influenzae species. A truncated protein has a significantly higher
CC amount of recovery than a full-length protein.
XX
SQ Sequence 3300 BP; 1099 A; 606 C; 819 G; 776 T; 0 other;
XX
Query Match 20.7%; Score 25.4; DB 21; Length 3300;
Best Local Similarity 61.2%; Pred. No. 27;
Matches 41; Conservative 0; Mismatches 26; Indels 0; Gaps 0;
OY 57 agggctgcaactgtaagttcacatgtaagtgatactatctcaattatcca 116
DB 154 agtgcctctgtctacgagtaagttgtaagttatgagcgatacgaataatcaatgca 213
OY 117 cacaca 123
DB 214 gccacata 220
XX
RESULT 21
AAH83942/C
ID AAH83942 standard; DNA; 649 BP.
XX
AC AAH83942;
XX
DT 25-SEP-2001 (first entry)
XX
DE Callithrix jacchus olfactory receptor encoding gene 10.
XX
KW Olfactory receptor; primate; mouse; human; food processing industry;
KW aromas; perfumery; toxic substance; ds.
KW Callithrix jacchus.
XX
OS
XX WO200146262-A2.
PN

```

PD 28-JUN-2001.
XX
XX 22-DEC-2000; 2000WO-IB02017.
XX
XX 22-DEC-1999; 99US-0171746.
PR 21-DEC-2000; 2000US-0747155.
XX
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Rouquier S, Giorgi D;
XX
XX WPI: 2001-381911/40.
DR P-PSDB: AAG98470.
XX
XX Nucleic acids encoding primate and murine olfactory receptors, useful
PT for analysis odours e.g. in food processing and perfumery -
XX
XX Claim 1; Page 150-151; 482pp; English.
XX
XX The invention relates to olfactory receptors (AAG98432-AAG98609) and the
CC genes encoding them (AAH83879-AAH84131) including pseudogenes of 10
CC primate species, mouse and human. The nucleic acids and receptors may be
CC used in the food processing industry (e.g. for the detection of aromas,
CC quality control and sample analysis), in perfumery (e.g. for the analysis
CC or comparison of perfumes) and in the environment (e.g. for the detection
CC of toxic substances and/or trapping of odours).
XX
XX Sequence 649 BP; 126 A; 192 C; 139 G; 192 T; 0 other;
SQ
Query Match 20.5%; Score 25.2; DB 22; Length 649;
Best Local Similarity 60.0%; Pred. No. 18; Indels 0; Gaps 0;
Matches 42; Conservative 0; Mismatches 28;
QY 1 caggatgtgaatgtgagatgccttacaacctgagccaatgycgaatgcgaagg 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 271 CAGGGAATGTAAGTGGAGATGCTGCACAGAACCCATATGATGAGTGGCAATGCCGGGG 212
QY 61 ctgcaactgac 70
    ||| ||| |||
DB 211 GTTCATAGCC 202
XX
XX RESULT 22
XX ID AAI60262 standard; CDNA; 3684 BP.
XX
XX AAI60262;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 4251.
XX
XX Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
XX Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
XX

```

```

PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AU, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
DR P-PSDB: AAM41106.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 4251; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cyostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC activation/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 3684 BP; 970 A; 818 C; 968 G; 928 T; 0 other;
SQ
Query Match 20.5%; Score 25.2; DB 22; Length 3684;
Best Local Similarity 52.9%; Pred. No. 33;
Matches 54; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
QY 21 tgccttacaacctgtgacctaatgycgaatgcgagggctgcaactgactgtaagttc 80
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3123 tgcattgcaccccccattgctctgtccccaagctaccacttaagtttaacttttt 3182
QY 81 cacatgtaagtgatcatcttcaatttcaatttccacacac 122
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3183 actagtcaaatagatcttcaactcaataatattccacacac 3224
XX
XX RESULT 23
XX ID AAT33537 standard; DNA; 12412 BP.
XX
XX AAT33537;
XX
XX 15-FEB-1998 (first entry)
XX
XX BCG deletion region 3 and flanking sequences.
XX
XX BCG delta 3; virulence; avirulence; attenuation; gene deletion;
KW mycobacteria; vaccine; infection; marker; ss.
XX
XX Mycobacterium bovis strain BCG.
XX
XX Key Location/Qualifiers
XX misc_feature 1406..10673
XX /tag= a
XX /note= "BCG delta 1 deletion region"
XX
XX WO9625519-A1.
XX

```

PD 22-AUG-1996.
XX
XX 15-FEB-1996; 96WO-US01938.
XX
PR 17-FEB-1995; 95US-0390878.
XX
PA (PATH-) PATHOGENESIS CORP.
XX
PI Mahairas GG, Stover CK;
XX
DR WPI; 1996-393419/39.
XX
XX Detecting markers for avirulence in *Mycobacterium* - used in
PT production of vaccines against bacterial infection, and to detect
PT bacterial infection
XX
XX Example 1; Fig 3; 66pp; English.
XX
XX This DNA sequence comprises *Mycobacterium bovis* BCG deletion
CC sequence BCGdelta3. A specific genetic deletion of this region
CC results in an avirulence phenotype of the mycobacterium. 2 other
CC deletion regions (see AAI33535 and AAI33536) have also been detected.
CC Identification involved screening a BCG cosmid library with a
CC radiolabeled probe obtained following DNA subtraction between
CC virulent *Mycobacterium tuberculosis* H37Rv and avirulent BCG.
CC The deletions provide useful markers for the identification of an
CC avirulent, or a virulent, mycobacterial phenotype. Determination
CC of avirulence requires the detection of the presence or absence of
CC the deletion; the deletions are detected either by detecting the
CC presence or absence of deletion junctions (see AAI33538-46), or by
CC detecting the presence or absence of the sequences contained within
CC the deletion. Deletion polypeptides are used as components of
CC immunological assays and in vaccines.
XX
SQ Sequence 12412 BP; 2146 A; 4273 C; 3876 G; 1917 T; 200 other;

Query Match 20.5%; Score 25.2; DB 17; Length 12412;
Best Local Similarity 66.7%; Pred. No. 50;
Matches 36; Conservative 0; Mismatches 18; Indels 0; Gaps 0;

QY 17 gagatgccttacacacctgatacctaattgacgaatgagggctgactgac 70
DB 5990 gacatgcttcaacgcgcgatgctgtaigcgcactgcaaacgcctgtgtgac 6043

RESULT 24
AAH32402/C
ID AAH32402 standard; DNA; 933 BP.
XX
AC AAH32402;
XX
DT 30-JUL-2001 (first entry)
XX
XX Human olfactory receptor polynucleotide, SEQ ID NO: 975.
DE
XX Human olfactory receptor; OR: primary scent determination;
KM secondary scent determination; polypeptide library; odour receptor;
KW scent profile; scent fingerprint; scent representation; ds.
XX
OS Homo sapiens.
XX
XX WO200127158-A2.
PN
XX 19-APR-2001.
PD
XX 06-OCT-2000; 2000WO-US27582.
PF
XX 08-OCT-1999; 99US-0158615.
PR
XX 24-FEB-2000; 2000US-0184809.
XX
PA (DIGI-) DIGISCENTS.
XX (YEDA) YEDA RES & DEV CO LTD.

XX
PI Bellenson J, Smith D, Lancet D, Glusman G, Fuchs T, Yanai I;
XX
DR WPI; 2001-290713/30.
XX
PT New polynucleotides which encode polypeptides involved in olfactory
PT sensation for identifying olfactory agonists and antagonists -
XX
PS Claim 8; Page 569; 1857pp; English.
XX
XX The present sequence is one of a number of isolated polynucleotides
CC which encode polypeptides involved in olfactory sensation. The
CC polynucleotides can be used in screening for olfactory agonists and
CC antagonists. The methods allow for the determination of primary
CC scents and the identification of the odour receptors used to detect
CC these primary scents. The methods also enable determination of
CC secondary scents and the identification of combinations of odour
CC receptors that are involved in detecting such secondary scents.
CC This enables the construction of a scent representation (also called
CC a scent fingerprint or scent profile), which may be used to re-create
CC and edit scents. Libraries of olfactory receptors are useful for
CC determining the interaction pattern of a composition with the receptors,
CC and can be used for determining differences in the olfactory faculties
CC of different individuals.
XX
SQ Sequence 933 BP; 206 A; 267 C; 187 G; 273 T; 0 other;

Query Match 20.3%; Score 25; DB 22; Length 933;
Best Local Similarity 61.5%; Pred. No. 24;
Matches 40; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 1 caggtatgtgaatgtgagatgcttaccacacctgactgactgacgaatgaggg 60
DB 474 CAGGGAATGTGAGTGGGAGATGCTGCAGAGAACCAATGATGATGATGCCATGG 415
QY 61 ctgca 65
DB 414 TTTC A 410

RESULT 25
AAAX04441/C
ID AAAX04441 standard; cDNA; 2284 BP.
XX
AC AAAX04441;
XX
DT 27-APR-1999 (first entry)
XX
DE S. frugiperda immunophilin FKBP46 cDNA.
XX
KW Immunophilin; moth; insect cell; nuclear; immunosuppression; drug;
KW transplant; tissue graft; ds.
XX
OS Spodoptera frugiperda.
XX
FH key Location/Qualifiers
FT CDS 160..1398
FT /*tag= a
FT /product= "immunophilin FKBP46"
FT 1764..1769
FT /*tag= b
FT polya_signal 1850..1855
FT /*tag= c
FT polya_signal 1868..1873
FT /*tag= d
FT polya_signal 2052..2057
FT /*tag= e
FT polya_signal 2237..2242
FT /*tag= f
XX
XX US5861498-A.
XX
XX

PD 19-JAN-1999.
 XX 31-OCT-1996; 96US-0741134.
 XX 01-NOV-1995; 95US-0007163.
 PR 31-OCT-1996; 96US-0741134.
 XX (UYJE-) UNIV JEFFERSON THOMAS.
 PA Alnemri ES, Fernandes-Alnemri T, Litwack G;
 PI WPI: 1999-130433/11.
 DR P-PSDB; AAM68010.
 XX
 XX Isolated immunophilin FKBP46 nucleic acids - useful for developing
 PT products for the study and identification of immunosuppressive
 PT agents for treating e.g. transplant and tissue graft patients
 XX
 PS Claim 2; Fig 1; 21pp; English.
 XX
 CC This sequence represents the cDNA encoding the immunophilin FKBP46 from
 CC the moth *Spodoptera frugiperda*. FKBP46 is a novel insect cell nuclear
 CC immunophilin, which can be used to study and identify additional
 CC immunosuppressive drugs that bind to it. Such immunosuppressive agents
 CC can be used in treating transplant and tissue graft patients.
 CC
 XX
 SQ Sequence 2284 BP; 803 A; 373 C; 495 G; 613 T; 0 other;
 Query Match 20.3%; Score 25; DB 20; Length 2284;
 Best Local Similarity 52.4%; Pred. No. 33; Mismatches 50; Indels 0; Gaps 0;
 Matches 55; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
 QY 4 gttatgttaagtgaagatgccttaacacctgatacctaagcgcaatgcgaggtcgtg 63
 Db 1755 GAATTGATGATCATCATGTCGCAATAATCCTTATGACATTAAGACAACTTATGCTG 1696
 QY 64 cactgaactggtgaagtttcacatgtaagtgatcattatcctaatt 108
 Db 1695 TAGACACCAATGTGTATCTGCAATATTTTACTGATGTTCAAT 1651
 RESULT 26
 AAX23518/c
 ID AAX23518 standard; DNA; 49998 BP.
 XX
 AC AAX23518;
 XX
 DT 23-JUN-1999 (first entry)
 XX
 DE Human kidney aminopeptidase P genomic DNA fragment 2.
 XX
 KW Aminopeptidase; human; AMP; gene therapy; treatment; AMP-deficiency;
 KW prenatal diagnosis; angiodema; antihypertensive agent; atherosclerosis;
 KW arterial stenosis; industrial protein feed; malabsorption syndrome;
 KW proteolaceous waste degradation; additive; immunohistochemistry; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO9911799-A2.
 XX
 PD 11-MAR-1999.
 XX
 PF 02-SEP-1998; 98MO-US18426.
 XX
 PR 02-SEP-1997; 97US-0057854.
 XX
 PA (MEDT-) MEDICAL COLLEGE GEORGIA RES INST.
 XX
 PI Ryan JW, Sprinkle TJC, Venema RC;
 XX
 DR WPI: 1999-205193/17.
 XX

PT Nucleic acid encoding human aminopeptidase P
 XX
 PS Claim 13; Page 109-139; 201pp; English.
 XX
 CC This invention describes the isolation of a novel human aminopeptidase P
 CC (AMP). This protein is used to produce recombinant AMP and can be used
 CC for gene therapy for treating AMP-deficiency conditions. Its fragments
 CC are used as primers and probes to identify patients with homozygous and
 CC heterozygous AMP deficiency, including prenatal diagnosis. (patients
 CC defective in AMP are at risk of developing angiodema if treated with
 CC angiotensin-converting enzyme inhibitors), also as antisense inhibitors
 CC in cases of excessive AMP expression. The product of the invention is
 CC also used to identify AMP-expressing sequences in other animals and to
 CC generate transgenic animals, and comparisons of genomic sequences are
 CC used to detect mutations. AMP inhibitors are potentially useful as
 CC antihypertensive agents and to prevent or treat arterial (re)stenosis
 CC or atherosclerosis. The structure of AMP is used to design synthetic
 CC substrates, e.g. for use in AMP assays. AMP, which hydrolyzes N-terminal
 CC imido bonds, can be used to degrade industrial protein feeds to free
 CC amino acids, to degrade proteinaceous wastes, as additives in enzyme
 CC formulations used to treat malabsorption syndrome and for studying its
 CC biological role. Antibodies against AMP are used in immunohistochemical
 CC methods to study AMP distribution.
 CC
 XX
 SQ Sequence 49998 BP; 12605 A; 11725 C; 11351 G; 14317 T; 0 other;
 Query Match 20.3%; Score 25; DB 20; Length 49998;
 Best Local Similarity 55.1%; Pred. No. 95;
 Matches 49; Conservative 0; Mismatches 40; Indels 0; Gaps 0;
 QY 11 aatgtgatagtccttaacacctgatacctaagcgcaatgcgaggtcgtcactgac 70
 Db 42441 AGATGTCACAGGCGCTCAGCATTTATATCACCACAGGTAGCCAGCTGGGGAAGAGCGC 42382
 QY 71 tggtaagttccacatgtaagtgatcactt 99
 Db 42381 AGGTCAAAATGCTCTAGTATGGGCCCTT 42353
 RESULT 27
 AAH13128/c
 ID AAH13128 standard; cDNA; 528 BP.
 XX
 AC AAH13128;
 XX
 DT 26-JUN-2001 (first entry)
 XX
 DE Human cDNA clone (3'-primer) SEQ ID NO:9963.
 XX
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
 XX
 KW Homo sapiens.
 OS
 PN EP1074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI: 2001-318749/34.
 XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 3; SEQ ID 9963; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 528 BP; 123 A; 126 C; 127 G; 132 T; 20 other;

Query Match 20.2%; Score 24.8; DB 22; Length 528;
Best local Similarity 60.3%; Pred. No. 23;
Matches 41; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 31 ccctgtgacctaatgacgaatgcagggctgactgactgtaagtacacatgtaag 90
DB 95 CCTGGAAGAGCTGTGCTCAGCAGCGGCTTCAAGGGATGGAACTTTATTTCAAA 36
QY 91 gtgtact 98
DB 35 TAAACTACT 28

RESULT 28
AAC39501/c
ID AAC39501 standard; DNA; 1448 BP.
XX
AC AAC39501;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 24864.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KM protein identification; signal transduction pathway;
KM metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138547.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148584.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 01-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159383.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161923.
PR 29-OCT-1999; 99US-0162142.

Query Match 20.2%; Score 24.8; DB 21; Length 1448;
Best Local Similarity 60.3%; Pred. No. 33;
Matches 41; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Oy 6 attgtaaatgtagatgcttaccacccctgacctaattgacgaatgagagctcga 65
Db 1341 ACTGAGACTGATGCTCCCTACACACTCTACACAGCTAATGATGATCCATACACTAGA 1282
Oy 66 ctgactcg 73
Db 1281 ATTGCTGCG 1274

RESULT 29
AAF23414/C
ID AAF23414 standard; DNA; 1782 BP.
XX
AC AAF23414;
XX
DT 20-MAR-2001 (first entry)
XX
DE Human SEC5 DNA sequence clone 3581980.0.30 SEQ ID 9.
XX

SEC5; secreted protein; cancer; angiogenesis; wound healing;
immune disorder; neurodegenerative disease; allergic reaction;
respiratory problem; organ transplantation; contraceptive; human;
chromosome 8; proliferative disorder; ds.

OS Homo sapiens.

PN WO200070046-A2.

PD 23-NOV-2000.

PF 12-MAY-2000; 2000WO-US13291.

PR 14-MAY-1999; 99US-0134315.

PR 12-JAN-2000; 2000US-0175744.

PR 10-MAR-2000; 2000US-0188274.

XX (CURA-) CURAGEN CORP.
XX Shinkets RA, Fernandes E, Boldog F;
PI


```
XX WPI: 2001-025020/03.
DR P-PSDB: AAB49653.
XX
PT New SECX polypeptides and nucleic acids useful for treating or
PT preventing cancer, other disorders related to angiogenesis,
PT neurodegenerative diseases, autoimmune disorders and allergic reactions
PT
XX
PS Claim 9; Page 20-21; 132pp; English.
XX
CC polynucleotide sequences AAF23410 - AAF23419 encode secreted SECX
CC proteins AAB49649 - AAB49658. Sequences AAF23420 - AAF23450 represent
CC primers and probes used in the isolation and characterization of the SECX
CC DNA sequences of the invention. The new polypeptides and nucleic acids
CC can be used in screening assays, detection assays, preventive or
CC predictive medicine, therapeutic and prophylactic treatment, and
CC pharmacogenomics. Specifically, the SECX polypeptides and nucleic acids
CC are useful for treating cancer; other disorders related to angiogenesis
CC e.g. abnormal wound healing, psoriasis; neurodegenerative diseases;
CC immune disorders; liver cirrhosis; benign tumours; fibrocystic conditions
CC and tissue hypertrophy (e.g. benign prostatic hypertrophy); allergic
CC reactions and conditions such as asthma and other respiratory problems;
CC as well as in treating or preventing diseases associated with organ
CC transplantation, atherosclerosis-associated diseases or disorders. The
CC polypeptides can also be used for bone, cartilage, tendon, ligament
CC and/or tissue growth or regeneration, wound healing, tissue repair and
CC replacement, gut protection or regeneration, as a contraceptive, to
CC inhibit thromboses, infections caused by bacteria, virus, fungi and other
CC parasites, and as a vaccine. SECX antibodies may be used to isolate or
CC detect SECX proteins, monitor protein level in tissue as part of a
CC clinical testing procedure, treat proliferative disorders including
CC tumours and benign hyperplasias.
XX
SQ Sequence 1782 BP; 521 A; 416 C; 397 G; 448 T; 0 other;

Query Match          20.2%; Score 24.8; DB 22; Length 1782;
Best Local Similarity 53.0%; Pred. No. 35;
Matches 53; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 14 tgtgatgccttacacacctgatgacctaatgagcgcaatgcgagggctgcacatgctgg 73
   ||| ||| | | | | | | | | | | | | | | | | |
DB 629 TCTTTATGTCAATGAATGATCTAATGCCACATGTCTGACTGGGTAGTAGCATGATATTC 570
   ||| ||| | | | | | | | | | | | | | | | | |
QY 74 taagttcacatgtaagtgatgactatcttcaattatt 113
   | | | | | | | | | | | | | | | | | | | |
DB 569 TCTGGCTCATTTAGAAAGGATGATCATCTTCAAGATGTT 530
   | | | | | | | | | | | | | | | | | | | |

RESULT 30
AAH16592
ID AAH16592 standard; cDNA; 1970 BP.
XX
AC AAH16592;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human cDNA sequence SEQ ID NO:15679.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
XX
PR 27-AUG-1999; 99JP-0300253.
XX
PR 11-JAN-2000; 2000JP-0118776.
```

```
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI: 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 15679; 2537pp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SQ Sequence 1970 BP; 434 A; 538 C; 552 G; 446 T; 0 other;

Query Match          20.2%; Score 24.8; DB 22; Length 1970;
Best Local Similarity 60.3%; Pred. No. 37;
Matches 41; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 31 cccctgagcactatggcgcaatgcgaggtgcacatgtaagttcacatgtaag 90
   || |||| | | | | | | | | | | | | | | | |
DB 1876 cctcgagacgctgtgctccacacgagcgactcaaggaagtggaaactttattcaaaa 1935
   || | | | | | | | | | | | | | | | | | | |

QY 91 gtgatact 98
   |||||
DB 1936 taataact 1943

RESULT 31
AAO80535
ID AAO80535 standard; cDNA to mRNA; 2064 BP.
XX
AC AAO80535;
XX
DT 31-MAY-1996 (first entry)
XX
DE Parolichthys olivaceus transglutaminase gene.
XX
DE Transglutaminase; fish; recombinant production; Pagrus major;
KW Theragra chalcogramma; Parolichthys olivaceus; Oncorhynchus keta; ds.
XX
OS Parolichthys olivaceus.
XX
PN JP07023787-A.
XX
```

PD 27-JAN-1995.
 XX
 XX 13-JUL-1993; 93JP-0172998.
 XX
 PR 13-JUL-1993; 93JP-0172998.
 XX
 PA (AJIN) AJINOMOTO KK.
 XX
 DR WPI; 1995-100948/14.
 DR P-PSDB; AAR65330.
 XX
 PT A fish-originated trans:glutaminase gene - can be produced in large
 PT amounts at low economic cost.
 XX
 PS Claim 12; Page 53; 94pp; Japanese.
 XX
 CC The DNA encodes a protein, isolated from Paralichthys olivaceus, which
 CC has transglutaminase (TG) activity. The TG's of the invention can be
 CC isolated from Pagrus major, Theragra chalcogramma, Paralichthys
 CC olivaceus or Oncorhynchus keta. Fish-originated TG's are economical
 CC to produce in large quantities.
 XX
 SQ Sequence 2064 BP; 548 A; 482 C; 586 G; 448 T; 0 other;

Query Match 20.2%; Score 24.8; DB 16; Length 2064;
 Best Local Similarity 57.9%; Pred. No. 37;
 Matches 44; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY 8 tgaatgtgagatgccttacacacctgacctaattgagcgcaatgcgaggtcgact 67
 ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| |||
 Db 633 tgcctatgtcagcagatgacacgcgcgtggtgaacctaacggtgacaggggtgtgct 692
 OY 68 gactggtgaagtttcac 83
 ||||| ||| |||
 Db 693 gactggtcagtcgacac 708

RESULT 32
 AA080533
 ID AA080533 standard; cDNA to mRNA; 2064 BP.
 XX
 AC AA080533;
 XX
 DT 31-MAY-1996 (first entry)
 XX
 DE Paralichthys olivaceus transglutaminase gene.
 XX
 KW transglutaminase; fish; recombinant production; Pagrus major;
 KW Theragra chalcogramma; Paralichthys olivaceus; Oncorhynchus keta; ds.
 XX
 OS Paralichthys olivaceus.
 XX
 PN JF07023787-A.
 XX
 PD 27-JAN-1995.
 XX
 PE 13-JUL-1993; 93JP-0172998.
 XX
 PR 13-JUL-1993; 93JP-0172998.
 XX
 PA (AJIN) AJINOMOTO KK.
 XX
 DR WPI; 1995-100948/14.
 DR P-PSDB; AAR65330.
 XX
 PT A fish-originated trans:glutaminase gene - can be produced in large
 PT amounts at low economic cost.
 XX
 PS Claim 10; Page 49; 94pp; Japanese.
 XX
 CC The DNA encodes a protein, isolated from Paralichthys olivaceus, which
 CC has transglutaminase (TG) activity. The TG's of the invention can be

CC Isolated from Pagrus major, Theragra chalcogramma, Paralichthys
 CC olivaceus or Oncorhynchus keta. Fish-originated TG's are economical
 CC to produce in large quantities.
 XX
 SQ Sequence 2064 BP; 548 A; 482 C; 585 G; 449 T; 0 other;

Query Match 20.2%; Score 24.8; DB 16; Length 2064;
 Best Local Similarity 57.9%; Pred. No. 37;
 Matches 44; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

OY 8 tgaatgtgagatgccttacacacctgacctaattgagcgcaatgcgaggtcgact 67
 ||| ||||| ||| ||| ||||| ||| ||| ||||| ||| |||
 Db 633 tgcctatgtcagcagatgacacgcgcgtggtgaacctaacggtgacaggggtgtgct 692
 OY 68 gactggtgaagtttcac 83
 ||||| ||| |||
 Db 693 gactggtcagtcgacac 708

RESULT 33
 AA046026
 ID AA046026 standard; cDNA to mRNA; 2148 BP.
 XX

AC AA046026;
 XX
 DT 21-FEB-1994 (first entry)
 XX
 DE Transglutaminase (fish liver) (T1882).
 XX

KW Transglutaminase; TGase; fish; Pagrus major; Alaska pollack;
 KW Theragra chalcogramma; Paralichthys olivaceus; paste; Kamaboko;
 KW gelatinous food; cosmetic; ss.
 XX
 OS Paralichthys olivaceus.
 XX

PH Key Location/Qualifiers
 FT 26..2092
 FT CDS /*tag= a

PN EP555649-A.
 XX
 PD 18-AUG-1993.
 XX
 PE 14-JAN-1993; 93EP-0100488.
 XX
 PR 14-JAN-1992; 92JP-0005166.
 PR 27-JUL-1992; 92JP-0198003.
 PR 08-DEC-1992; 92JP-0328010.
 XX
 PA (AJIN) AJINOMOTO KK.
 XX
 PI Matsui H, Motoki M, Nagase K, Nakanishi K, Yasueda H;
 PI WPI; 1993-259983/33.
 XX
 DR P-PSDB; AAR39519.
 XX
 PT New gene encoding trans:glutaminase from fish - for the mass
 PT prodn. of trans:glutaminase for use in gelatinous food, cosmetics
 PT and chemicals
 XX
 PS Claim 1; Page 42-44; 85pp; English.
 XX
 CC The difference between the sequences of AA046025 and AA054878 resides
 CC in the 1882 th base. The change does not alter the translated
 CC prod.
 CC Transglutaminase is used in the prodn. of gelatinous food prods.,
 CC cosmetics and chemicals. The DNA fragment enables the mass prodn.
 CC of transglutaminase which is cheap and efficient.
 XX
 SQ Sequence 2148 BP; 578 A; 496 C; 610 G; 464 T; 0 other;

AC	AAQ80534;
XX	
DT	31-MAY-1996 (first entry)

```

XX Paralicthys olivaceus transglutaminase gene.
DE transglutaminase; fish; recombinant production; Pagrus major;
KW Theragra chalcogramma; Paralicthys olivaceus; Oncorhynchus keta; ds.
OS Paralicthys olivaceus.
XX
FH Key Location/Qualifiers
FT CDS 26..2092
FT /tag= a
FT /product= transglutaminase
XX
XX JP07023787-A.
XX
XX 27-JAN-1995.
XX
XX 13-JUL-1993; 933P-0172998.
XX
XX 13-JUL-1993; 933P-0172998.
XX
XX (AJIN ) AJINOMOTO KK.
XX
XX WPI; 1995-100948/14.
XX
XX P-PSDB; AAR65331.
XX
XX A fish-originated trans:glutaminase gene - can be produced in large
XX amounts at low economic cost.
XX
XX Disclosure; Page 50-53; 94pp; Japanese.
XX
XX The DNA encodes a protein, isolated from Paralicthys olivaceus, which
XX has transglutaminase (TG) activity. The TG's of the invention can be
XX isolated from Pagrus major, Theragra chalcogramma, Paralicthys
XX olivaceus or Oncorhynchus keta. Fish-originated TG's are economical
XX to produce in large quantities.
XX
XX Sequence 2148 BP; 578 A; 496 C; 610 G; 464 T; 0 other;
XX
Query Match 20.2%; Score 24.8; DB 16; Length 2148;
Best Local Similarity 57.9%; Pred. No. 38;
Matches 44; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
QY 8 tgaatgtgagatgcctacaaccctgatagcctaagcgcaatgcgaggtcgcac 67
Db 661 tgcctatgcagcagatgatactcgtgtaacttaacgagcgaggtgtgtgc 720
QY 68 gactgtgaagttctac 83
Db 721 gactgtgtaagttgcac 736

```

RESULT 37

AAC59911

ID AAC59911 standard; cDNA; 452 BP.

XX

AC AAC59911;

XX

DT 30-JAN-2001 (first entry)

XX

DE Human secreted protein cDNA sequence #5.

XX

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;

KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;

KW vunerary; anticonvulsant; antibacterial; antifungal; antiparasitic;

KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;

KW neurological disease; infection; human; secreted protein; ss.

XX

OS Homo sapiens.

XX

XX WO200055198-A1.

XX

```

PD 21-SEP-2000.
XX
XX 09-MAR-2000; 2000WO-US06012.
XX
XX 12-MAR-1999; 99US-0124093.
XX
XX 23-NOV-1999; 99US-0166989.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM, Komatsoulis G;
XX
XX WPI; 2000-587520/55.
XX
XX P-PSDB; AAB28706.
XX
XX Human secreted proteins and the nucleic acids that encode them, useful
XX in gene therapy protocols and recombinant nucleic acid based procedures
XX
XX Claim 1; Page 315; 391pp; English.
XX
XX The invention relate to the isolation of genes AAC59907-C59956 encoding
XX 50 human secreted proteins AAB28702-B28751. The genes can be used to
XX generate fusion proteins by linking to the gene for the human
XX immunoglobulin G Fc portion for increasing the stability of
XX the fusion protein as compared to the human protein only. The genes and
XX proteins are useful for preventing, ameliorating or treating medical
XX conditions, e.g. by protein or gene therapy. The genes are isolated
XX from a range of human tissues disclosed in the specification. The
XX nucleic acids, proteins, antibodies and (ant)agonists are useful in
XX the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
XX and ovarian cancer, and other cancers of the adrenal gland, bone, bone
XX marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
XX (b) immune disorders e.g. Addison's disease, allergies, autoimmune
XX haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
XX disease, multiple sclerosis, rheumatoid arthritis and ulcerative
XX colitis; (c) cardiovascular disorders such as myocardial ischaemias; (d)
XX wound healing; (e) neurological diseases e.g. cerebral anoxia and
XX epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
XX and parasitic infections.
XX
XX Sequence 452 BP; 116 A; 104 C; 101 G; 128 T; 3 other;
XX
Query Match 20.0%; Score 24.6; DB 21; Length 452;
Best Local Similarity 65.5%; Pred. No. 26;
Matches 36; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
QY 67 tgactgtgaagtttcacatgtaagtgatactatcttaatttccacaaa 121
Db 209 tcactgttcagtgctactgttggtaacatacaatcaacgltgtctccaaa 263

```

RESULT 38

AAR26556

ID AAR26556 standard; DNA; 625 BP.

XX

AC AAR26556;

XX

DT 27-MAR-2001 (first entry)

XX

DE DNA encoding human secreted protein #10.

XX

XX Secreted protein; gene therapy; vaccine; cancer; leukemia;

KW autoimmune disease; allergy; inflammation; graft rejection;

KW hyperproliferation; cardiovascular; infection; ss.

XX

OS Homo sapiens.

XX

XX WO200076531-A1.

XX

PD 21-DEC-2000.

XX

PF 01-JUN-2000; 2000WO-US51317.

Query Match	20.0%	Score 24.6	DB 22	Length 625
Best Local Similarity	65.5%	Pred. No. 29		
Matches	36	Conservative	0	Mismatches 19; Indels 0; Gaps 0

RESULT	39
AAN71405/C	
ID	AAN71405 standard; DNA; 1864 BP

DE Sequence of ANS-1 which increases transformation efficiency.

OS *Mucor miehei*.

PD 25-MAR-1987.

PR 07-JUL-1986;

XX
XX

XX
XX

PT New DNA sequences for expressing polypeptide in filamentous fungi
PT - with secretion of prod. from the cells, and new vectors and
PT transformed fungi

PS Example; Fig 13; 45pp; English.

Query Match	20.0%	Score 24.6	DB 8	Length 1864
Best Local Similarity	48.6%	Pred. No. 43		
Matches 54	Conservative 0	Mismatches 57	Indels 0	Gaps 0

Qy 62 cgacgcactgtgaagttcacatgtaagtgtacctattctcaatttat 112
| | | | | | | | | | | | | | |
Db 339 TTAATATTATATTAANNATAAATATATNGGAAGNTACTCTATTTAT 289

```

RESULT  40
AAH62796/c
ID  AAH62796 standard; DNA; 2946 BP.

```

Shrimp white spot Bacilliform virus (WSBV) gene 107.

KW Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;

KW transgenic viral resistant shrimp; ds.

OS White spot syndrome virus.

PN WO200138351-A2

PD 31-MAY-2001

PF 08-NOV-2000; 2000WO-US28888.

PR 24-NOV-1999; 99CN-0124717.

PA (PENY-) PE CORP NY.

PA (SINO-) SINOGENOMAX CO LTD.

PI Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;

DR WPI; 2001-355877/

PT Primary nucleotide sequence of the shrimp white spot Bacilliform virus
PT (MSBV), useful for producing viral polypeptides that can be used to
PT screen for agents that are useful for treating MSBV infection -

PS Claim 4; Figure 3; 626pp; English.

CC The invention provides the primary nucleotide sequence of the MSV genome
CC (AAH62689). Predicted transcript sequences (AAH62689-AAH62839) and
CC encoded proteins (AAC684910-AA685011) and oligonucleotide sequences
CC (AAH62840-63160) suitable for use as primers or probes. The nucleic acid
CC molecules and proteins of the invention are useful for diagnosis and
CC monitoring viral infection. In screens for antiviral agents and for
CC monitoring viral gene expression or activity during a treatment regimen.
CC The nucleic acid molecules are also useful as antisense constructs to
CC control viraemia expression in infected cells and tissues and to create

CC transgenic viral resistant shrimp.

XX Sequence 2946 BP, 979 A, 566 C, 665 G, 736 T, 0 other:

Query Match 20.0%; Score 24.6; DB 22; Length 2946;
Best Local Similarity 70.2%; Pred. No. 50;
Matches 33; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 69 actggaattcacatgaagtgatactatctcaattatcc 115
Db 1549 ACTCGATTTCACATTCATTCAACTTACCTCACTCTCTCC 1503

RESULT 41

AAH81775
ID AAH81775 standard; DNA; 3787 BP.

XX AAH81775;

XX 21-SEP-2001 (first entry)

XX Human differential transcription-associated cDNA SEQ ID 284.

XX Differential transcription; human; rat; tumour cell; cytostatic;

XX Ras modulator; Class II tumour suppressor gene; gene therapy; ss.

XX Homo sapiens.

XX W0200157058-A2.

XX 09-AUG-2001.

XX 31-JAN-2001; 2001WO-EP01003.

XX 31-JAN-2000; 2000DE-1004102.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX Rosenthal A, Hinzmann B, Schaefer R, Zuber J, Tcherntsa O;

XX Grips M, Hellriegel M, Schmitz A, Sers C;

XX WPI; 2001-483415/52.

XX Nucleic acids differentially expressed between tumor and normal cells,
XX useful for diagnosis or therapy of tumors and for screening active
XX agents

XX Disclosure; Page 438-439; 579pp; German.

XX This invention describes a nucleic acid (I) with differential expression
XX between tumour and normal cells and which has cytostatic activity. (I)
XX work as modulators of Ras activity by inducing expression of tumour
XX suppressor genes. (I), and polypeptides encoded by them, are useful as
XX targets for diagnosis or therapy and in screening to determine the
XX effects of an active compound (potential pharmaceutical) on a cell line,
XX particularly for diagnosis and treatment of tumors, especially by
XX modulating expression of (I) (by gene therapy, antisense RNA or ribozyme
XX methods) or by modulating the amount and/or location of (I)-encoded
XX polypeptides (by administration of the polypeptide or its activator,
XX antibody (optionally as a conjugate) or inhibitor). The method allows
XX identification of many Class II tumour suppressor genes (i.e. genes that
XX are not primary targets for tumour-initiating mutations).
XX AAH81492-AAH82376 represent the human and rat derived nucleic acid
XX fragments described in the method of the invention.

XX Sequence 3787 BP; 898 A; 1210 C; 1073 G; 606 T; 0 other;

Query Match 20.0%; Score 24.6; DB 22; Length 3787;
Best Local Similarity 59.2%; Pred. No. 54;
Matches 42; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 4 gtatgttaatgtgagatgcttacacccctgatgacccaatggcgcaatgagagctg 63
Db 372 gtatgcgtctgccgcgtccctacagcagcttacccgttatgcatgagtcgacctg 431

QY 64 cactgactggt 74
Db 432 caagagactggt 442

RESULT 42

AAH57351
ID AAH57351 standard; DNA; 68940 BP.

XX AAH57351;

XX 16-AUG-1999 (first entry)

XX Human chromosome 6q27 tumour suppressor gene.

XX Tumour suppressor gene; chromosome 6q27; cancer; therapy; human; D6S297;

XX breast cancer; ovarian cancer; lymphoid cancer; marker; D6S193; ss.

XX Homo sapiens.

XX key Location/Qualifiers

XX exon 414..549

XX polyA_site /tag= a complement (719..724)

XX repeat_region /tag= b 1126..1220

XX variation /tag= c "MIR repeat family"

XX variation /note= "replace(1128,T)"

XX variation /note= "putative polymorphism, C in 52b3, T in

XX variation /note= "putative polymorphism, A in 52b3, C in

XX variation /note= "putative polymorphism, A in 52b3, C in

XX variation /note= "putative polymorphism, C in 52b3, T in

XX variation /note= "putative polymorphism, C in 52b3, T in

XX variation /note= "putative polymorphism, C in 52b3, T in

XX variation /note= "putative polymorphism, C in 52b3, T in

XX variation /note= "putative polymorphism, C in 52b3, T in

XX variation /note= "putative polymorphism, C in 52b3, T in

XX variation /note= "putative polymorphism, C in 52b3, T in

XX variation /note= "putative polymorphism, C in 52b3, T in

XX variation /note= "putative polymorphism, C in 52b3, T in

XX variation /note= "putative polymorphism, C in 52b3, T in

XX variation /note= "putative polymorphism, C in 52b3, T in

XX variation /note= "putative polymorphism, C in 52b3, T in

XX variation /note= "putative polymorphism, C in 52b3, T in

XX variation /note= "putative polymorphism, C in 52b3, T in

XX variation /note= "putative polymorphism, C in 52b3, T in

XX variation /note= "putative polymorphism, C in 52b3, T in

```

FT      /*tag= s
FT      6072..6134
FT      /*tag= t
FT      complement (6395..6438)
FT      exon
FT      /*tag= u
FT      /note= "trapped exon"
FT      complement (6395..6456)
FT      /*tag= v
FT      complement (6887..6892)
FT      polyA_site
FT      /*tag= w
FT      complement (7237..7434)
FT      exon
FT      /*tag= x
FT      complement (8055..8165)
FT      exon
FT      /*tag= y
FT      complement (8414..8447)
FT      exon
FT      /*tag= z
FT      8530..8695
FT      /*tag= aa
FT      complement (8543..8593)
FT      exon
FT      /*tag= ab
FT      /note= "trapped exon"
FT      9351..9678
FT      exon
FT      /*tag= ac
FT      /note= "trapped exon"
FT      complement (9467..9583)
FT      exon
FT      /*tag= ad
FT      9627..9678
FT      exon
FT      /*tag= ae
FT      complement (9636..9685)
FT      promoter
FT      /*tag= af
FT      complement (10423..10461)
FT      exon
FT      /*tag= ag
FT      complement (10669..10761)
FT      exon
FT      /*tag= ah
FT      complement (10706..10761)
FT      exon
FT      /*tag= ai
FT      11067..11368
FT      repeat_region
FT      /*tag= aj
FT      /note= "Alu repeat family"
FT      11264..11315
FT      exon
FT      /*tag= ak
FT      complement (12270..12275)
FT      polyA_site
FT      /*tag= al
FT      complement (12641..12690)
FT      promoter
FT      /*tag= am
FT      13800..14074
FT      exon
FT      /*tag= an
FT      /note= "trapped exon"
FT      complement (14197..14400)
FT      exon
FT      /*tag= ao
FT      complement (14496..14601)
FT      repeat_region
FT      /*tag= ap
FT      /note= "L1 repeat family"
FT      complement (14607..14860)
FT      exon
FT      /*tag= aq
FT      /note= "trapped exon"
FT      14930..14979
FT      promoter
FT      /*tag= ar
FT      replace(15091..G)
FT      variation
FT      /*tag= as
FT      /note= "putative polymorphism, A in 52b3 and
FT      pdj84g15, G in 255h6"
FT      complement (15609..15614)
FT      polyA_site
FT      /*tag= at
FT      16174..16451
FT      repeat_region
FT      /*tag= au
FT      /note= "Alu repeat family"
FT      16509..16559
FT      exon
FT      /*tag= av
FT      complement (16867..16939)
FT      exon
FT      /*tag= aw
FT      /note= "trapped exon"
FT      16867..17013
FT      exon

```

```

FT      /*tag= ax
FT      /note= "trapped exon"
FT      17288..17352
FT      exon
FT      /*tag= ay
FT      complement (17427..17606)
FT      exon
FT      /*tag= az
FT      17647..17696
FT      promoter
FT      /*tag= ba
FT      complement (18367..18416)
FT      promoter
FT      /*tag= bb
FT      complement (19301..19334)
FT      exon
FT      /*tag= bc
FT      complement (19541..19683)
FT      repeat_region
FT      /*tag= bd
FT      /note= "MER20 repeat family"
FT      19688..19975
FT      repeat_region
FT      /*tag= be
FT      /note= "Alu repeat family"
FT      20108..20226
FT      exon
FT      /*tag= bf
FT      replace(20729..20767)
FT      variation
FT      /*tag= bg
FT      /note= "putative polymorphism (TAT13) in 52b3"
FT      complement (20770..20878)
FT      repeat_region
FT      /*tag= bh
FT      /note= "Alu repeat family"
FT      21292..21341
FT      promoter
FT      /*tag= bi
FT      21450..21499
FT      promoter
FT      /*tag= bj
FT      complement (22392..22441)
FT      promoter
FT      /*tag= bk
FT      22856..22956
FT      exon
FT      /*tag= bl
FT      complement (23196..23233)
FT      exon
FT      /*tag= bm
FT      complement (23288..23293)
FT      polyA_site
FT      /*tag= bn
FT      replace(24182..24186)
FT      variation
FT      /*tag= bo
FT      /note= "putative polymorphism (T15 in pdj84g15"
FT      complement (24244..24293)
FT      promoter
FT      /*tag= bp
FT      24266..24315
FT      promoter
FT      /*tag= bq
FT      complement(24600..24649)
FT      promoter
FT      /*tag= br
FT      30550..31665
FT      exon
FT      /*tag= bs
FT      30550..30863
FT      exon
FT      /*tag= bt
FT      32483..32488
FT      polyA_site
FT      /*tag= bu
FT      32483..32488
FT      repeat_region
FT      /*tag= bv
FT      33639..33870
FT      repeat_region
FT      /*tag= bw
FT      /note= "Alu repeat family"
FT      complement (33782..33867)
FT      exon
FT      /*tag= bx
FT      complement (34638..34996)
FT      repeat_region
FT      /*tag= by
FT      /note= "TIGER2 repeat family"
FT      complement (34879..34941)
FT      exon
FT      /*tag= bz

```

```

Query Match      20.0%; Score 24.6; DB 20; Length 68940;
Best Local Similarity 52.4%; Pred. No. 1.5e+02;
Matches 54; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
Oy      14  tgtgagatgccttacacacctgatacctaattgacgaatgcgaaggctgactgactg 73
          || | ||| ||| | | | | | | | | | | | | | | | | | | | | |
Db      48374 tgaataatgacttaagaatgatatagtcataatacatgcatgcggtgcctccaaagagatg 48433

```



```
FT CDS complement (38892..40163)
FT /tag= au
FT /product= "EvrS"
FT CDS complement (40216..40890)
FT /tag= av
FT /product= "EvrT"
FT RBS complement (40899..40902)
FT /tag= aw
FT CDS complement (40887..41576)
FT /tag= ax
FT /product= "EvrU"
FT CDS complement (41679..42707)
FT /tag= ay
FT /product= "EvrV"
FT RBS complement (42714..42717)
FT /tag= az
FT CDS complement (42810..43799)
FT /tag= ba
FT /product= "EvrW"
FT RBS complement (43807..43811)
FT /tag= bb
FT CDS complement (43799..44866)
FT /tag= bc
FT /product= "EvrX"
FT CDS complement (45014..45760)
FT /tag= bd
FT /product= "EvrY"
FT RBS complement (45767..45770)
FT /tag= be
FT CDS complement (45962..46714)
FT /tag= bf
FT /product= "EvrZ"
FT RBS complement (45952..45956)
FT /tag= bg
FT CDS complement (47156..49234)
FT /tag= bh
FT /product= "EvsA"
FT CDS 51627..52715
FT /tag= bi
FT /product= "EvsB"
FT RBS 51629..51622
FT /tag= bj
FT CDS 52889..53557
FT /tag= bk
FT /product= "EvsC"
FT CDS 53554..54207
FT /tag= bl
FT /product= "EvsD"
FT CDS complement (54362..55117)
FT /tag= bm
FT /product= "EvsE"
FT RBS complement (55125..55128)
FT /tag= bn
FT CDS complement (55135..56094)
FT /tag= bo
FT /product= "EvsF"
FT RBS complement (56100..56103)
FT /tag= bp
FT CDS complement (56184..56813)
FT /tag= bq
FT /product= "EvsG"
FT CDS 56961..58709
```

```
Query Match 20.0%; Score 24.6; DB 22; Length 109519;
Best Local Similarity 65.5%; Pred. No. 1.7e+02;
Matches 36; Conservative 0; Mismatches 19; Indels 0; Gaps 0;
```

```
OY 18 agatgcttaccacactgacgaatgcgaagcgtcactgactg 72
DB 76022 AGATGGGCTACGACTCGACGCCGATCCACCGCGAGTGGCAAGTGG 75968
```

RESULT 44

```
AACT5339/C
ID AAC75339 standard; cDNA; 294 BP.
XX
AC AAC75339;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF894 polynucleotide sequence SEQ ID NO:1787.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
vulnerable; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;
hypotensive; dermatological; immunosuppressive; antiinflammatory;
antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
neurodegenerative disorder; osteoarthritis; graft vs host disease;
cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
cholesterol ester storage; systemic lupus erythematosus; infection;
severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
bone damage; cartilage damage; antinflammatory disease; coagulation;
thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN M0200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000MO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shinkets RA, Leach M;
XX
DR MPI; 2000-602362/57.
XX
P-PSDB; AAB41130.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 1399; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerable;
CC antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiac; thrombolytic; coagulant; vasotrophic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antihypertensive; antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 294 BP; 96 A; 51 C; 68 G; 79 T; 0 other;
```

[illegible]

CC responsiveness of cancer to therapy). This is particularly for breast
CC cancer, oestrogen receptor-positive breast cancer, oestrogen receptor
CC negative breast cancer, lung cancer, and colon cancer.
XX
SQ Sequence 300 BP; 85 A; 72 C; 73 G; 70 T; 0 other;

[illegible]

Search completed: March 19, 2002, 10:21:41
Job time: 2850 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2002, 09:31:46 ; Search time 39 Seconds

(without alignments)
714.277 Million cell updates/sec

Title: US-09-684-016-48411

Perfect score: 123

Sequence: 1 caggtatgtaaatgtaga.....tcaatttaccacacaca 123

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCITUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25.6	20.8	5599	2	US-08-477-451-9
2	25.6	20.8	5599	2	US-08-477-451-13
3	25.6	20.8	19932	2	US-08-477-451-25
4	25.2	20.5	12412	1	US-08-390-878-18
5	25	20.3	2255	2	US-08-741-134-1
6	24.8	20.2	2064	1	US-08-164-839-30
7	24.8	20.2	2064	1	US-08-164-839-32
8	24.8	20.2	2064	1	US-08-583-799-30
9	24.8	20.2	2064	1	US-08-583-799-32
10	24.8	20.2	2148	1	US-08-164-839-69
11	24.8	20.2	2148	1	US-08-164-839-71
12	24.8	20.2	2148	1	US-08-583-799-69
13	24.8	20.2	2148	1	US-08-583-799-71
14	24.4	19.8	7720	4	US-09-318-448-5
15	24.2	19.7	1839	3	US-08-152-019A-44
16	24.2	19.7	1839	3	US-08-482-677-1
17	24.2	19.7	6642	1	US-08-727-034-5
18	24.2	19.7	6843	1	US-08-727-034-6
19	23.6	19.2	2329	1	US-08-253-785-2
20	23.4	19.0	562	2	US-08-975-316-53
21	23.4	19.0	1026	2	US-08-713-000-6
22	23.4	19.0	1026	2	US-08-975-316-6
23	23.4	19.0	1026	4	US-09-211-710-6
24	23	18.7	1959	4	US-09-201-664-1
25	23	18.7	2059	1	US-08-244-686-1
26	23	18.7	2277	1	US-08-369-796-7
27	23	18.7	2277	2	US-08-852-091-7

c 28	23	18.7	2277	2	US-08-820-754-7	Sequence 7, Appl
c 29	23	18.7	2277	3	US-08-956-652-7	Sequence 7, Appl
c 30	23	18.7	2277	3	US-08-956-865-7	Sequence 7, Appl
c 31	23	18.7	2277	3	US-08-948-547-7	Sequence 7, Appl
c 32	23	18.7	2277	5	PCR-US95-17025-7	Sequence 7, Appl
c 33	23	18.7	3464	4	US-09-318-448-30	Sequence 30, Appl
c 34	23	18.7	4233	4	US-09-056-105-27	Sequence 27, Appl
c 35	23	18.7	4440	3	US-07-792-600-1	Sequence 1, Appl
c 36	23	18.7	4440	3	US-09-157-021-1	Sequence 1, Appl
c 37	23	18.7	4440	3	US-09-156-842-1	Sequence 1, Appl
c 38	23	18.7	5433	3	US-09-157-021-35	Sequence 35, Appl
c 39	23	18.7	5433	3	US-09-156-842-35	Sequence 35, Appl
c 40	22.8	18.5	180	3	US-08-483-503A-2	Sequence 2, Appl
c 41	22.8	18.5	579	2	US-08-896-501A-5	Sequence 5, Appl
c 42	22.8	18.5	579	2	US-08-896-501A-5	Sequence 5, Appl
c 43	22.8	18.5	680	4	US-09-328-111-774	Sequence 774, App
c 44	22.8	18.5	911	4	US-08-943-731-81	Sequence 81, Appl
c 45	22.8	18.5	1120	3	US-08-884-324-2	Sequence 2, Appl

ALIGNMENTS

```
RESULT 1
US-08-477-451-9
; Sequence 9, Application US/08477451
; Patent No. 5928665
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori CagI Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,451
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: A35
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0335.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5599 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-477-451-9

Query Match      20.8%; Score 25.6; DB 2; Length 5599;
Best Local Similarity 51.8%; Pred. No. 5.5;
Matches 58; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 11 aaatggaatgccttacaccctgagtacatgagcgacatgagcgagctgacatgac 70
Db 4528 AATGCTTCACATCTGACATATTACAGCACCATTTTGCGCGCCCTTCGATGCTCTTG 4587
QY 71 tggtaagttcacatgtgaagtgatactatacttcaattatccacacaac 122
```

Db 4588 CGTTAAATTCCTGTTAATTGCTAAATATTAGATATTGTTCCACAGC 4639

RESULT 2

US-08-477-451-13/C

; Sequence 13, Application US/08477451
; Patent No. 5928865
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori CagI Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,451
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0335.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5599 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-477-451-13

Query Match 20.8%; Score 25.6; DB 2; Length 5599;
Best Local Similarity 51.8%; Pred. No. 5.5;
Matches 58; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

OY 11 aaatgtagatgaccttacaacctgatacctaatacgcgcaatgagggctgcactgac 70
Db 1072 AAATGCTTGCAATCTGACTATTAAACAGCACCAATTGGCGCCCTTGCGTGTG 1013

OY 71 tggtaagtctacatgtaagtgatctatcttcattatccacacaac 122
Db 1012 CGTTAAATTCCTGTTAATTGCTAAATATTAGATATTGTTCCACAGC 961

RESULT 3

US-08-477-451-25

; Sequence 25, Application US/08477451
; Patent No. 5928865
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori CagI Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,451
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0335.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-655-3542
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19932 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-477-451-25

Query Match 20.8%; Score 25.6; DB 2; Length 19932;
Best Local Similarity 51.8%; Pred. No. 8.9;
Matches 58; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

OY 11 aaatgtagatgaccttacaacctgatacctaatacgcgcaatgagggctgcactgac 70
Db 8085 AAATGCTTGCAATCTGACTATTAAACAGCACCAATTGGCGCCCTTGCGTGTGAG 8144

OY 71 tggtaagtctacatgtaagtgatctatcttcattatccacacaac 122
Db 8145 CGTTAAATTCCTGTTAATTGCTAAATATTAGATATTGTTCCACAGC 8196

RESULT 4

US-08-390-878-18

; Sequence 18, Application US/08390878
; Patent No. 5700683
; GENERAL INFORMATION:
; APPLICANT: Stover, Charles K.
; APPLICANT: Mahairas, Gregory G.
; TITLE OF INVENTION: VIRULENCE-ATTENUATING GENETIC DELETIONS
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza, Stewart Street Tower, 20th
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,878
; FILING DATE: 17-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 15371A-17
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/543/9600
; TELEFAX: 415/543/5043
; INFORMATION FOR SEQ ID NO: 18:

Query Match	20.3%	Score 25;	DB 2;	Length 2255;
Best Local Similarity	52.4%;	Pred. No. 6.5;		
Matches 55; Conservative	0;	Mismatches 50;	Indels 0;	Gaps 0;

Query Match	20.2%	Score 24.8;	DB 1;	Length 2064;
Best Local Similarity	57.9%	Pred. No. 7.4;		
Matches 44;	Conservative 0;	Mismatches 32;	Indels 0;	Gaps 0;
8 ttttaattgttgtagccttacaccctatgtaactaatgagcgaatgcgaggtgcact 67				

Db	633	TGTCATATCTACGACGAGATGATCACTGGCATGTGTAACCTCAACGCTACACGGGTTGCT	692
QY	68	gactcgttaagttaac	83
Db	693	GACTGCTCAGTCGGCAC	708

```

RESULT 7
US-08-164-839-32
Sequence 32, Application US/08164839
Patent No. 5514573
GENERAL INFORMATION:
APPLICANT: YASUEDA, HISASHI
APPLICANT: NAKANISHI, KAZUO
APPLICANT: MOTOKI, MASAO
APPLICANT: NAGASE, KAZUO
APPLICANT: MATSUI, HIROSHI
TITLE OF INVENTION: GENE ENCODING TRANSGLUTAMINASE DERIVED
FROM FISH
NUMBER OF SEQUENCES: 72
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLION, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/164,839
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/004,729
FILING DATE: 14-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Oblion, No. 5514573man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 10-599-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)412-3000
TELEFAX: (703)413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO.: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 2064 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Parachanna olivaceus
TISSUE TYPE: liver
FEATURE:
NAME/KEY: CDS
LOCATION: 1..2061
US-08-164-839-32

```

	Query Match	24.8%	Score 24.8;	DB 1,	Length 2064;
	Best Local Similarity	57.9%	Pred. No. 7.4;		
	Matches	44;	Conservative	0;	Mismatches 32; Indels 0; Gaps 0.
OY	8	tgtataatgtagagcgccttacaaccccgatgatgaacctaatggcgcgaatgcgagagtcgcact	67		
Db	633	TGTTATCTATGTACGCAGATGATCTACTGCCATGGTGAACCTCTAACGGGTGACAGGGGTGTCT	692		
OY	68	gactggttaagtltcaac	83		

```

      1
      2
      3
      4
      5
      6
      7
      8
      9
     10
     11
     12
     13
     14
     15
     16
     17
     18
     19
     20
     21
     22
     23
     24
     25
     26
     27
     28
     29
     30
     31
     32
     33
     34
     35
     36
     37
     38
     39
     40
     41
     42
     43
     44
     45
     46
     47
     48
     49
     50
     51
     52
     53
     54
     55
     56
     57
     58
     59
     60
     61
     62
     63
     64
     65
     66
     67
     68
     69
     70
     71
     72
     73
     74
     75
     76
     77
     78
     79
     80
     81
     82
     83
     84
     85
     86
     87
     88
     89
     90
     91
     92
     93
     94
     95
     96
     97
     98
     99
    100
    101
    102
    103
    104
    105
    106
    107
    108
    109
    110
    111
    112
    113
    114
    115
    116
    117
    118
    119
    120
    121
    122
    123
    124
    125
    126
    127
    128
    129
    130
    131
    132
    133
    134
    135
    136
    137
    138
    139
    140
    141
    142
    143
    144
    145
    146
    147
    148
    149
    150
    151
    152
    153
    154
    155
    156
    157
    158
    159
    160
    161
    162
    163
    164
    165
    166
    167
    168
    169
    170
    171
    172
    173
    174
    175
    176
    177
    178
    179
    180
    181
    182
    183
    184
    185
    186
    187
    188
    189
    190
    191
    192
    193
    194
    195
    196
    197
    198
    199
    200
    201
    202
    203
    204
    205
    206
    207
    208
    209
    210
    211
    212
    213
    214
    215
    216
    217
    218
    219
    220
    221
    222
    223
    224
    225
    226
    227
    228
    229
    230
    231
    232
    233
    234
    235
    236
    237
    238
    239
    240
    241
    242
    243
    244
    245
    246
    247
    248
    249
    250
    251
    252
    253
    254
    255
    256
    257
    258
    259
    260
    261
    262
    263
    264
    265
    266
    267
    268
    269
    270
    271
    272
    273
    274
    275
    276
    277
    278
    279
    280
    281
    282
    283
    284
    285
    286
    287
    288
    289
    290
    291
    292
    293
    294
    295
    296
    297
    298
    299
    300
    301
    302
    303
    304
    305
    306
    307
    308
    309
    310
    311
    312
    313
    314
    315
    316
    317
    318
    319
    320
    321
    322
    323
    324
    325
    326
    327
    328
    329
    330
    331
    332
    333
    334
    335
    336
    337
    338
    339
    340
    341
    342
    343
    344
    345
    346
    347
    348
    349
    350
    351
    352
    353
    354
    355
    356
    357
    358
    359
    360
    361
    362
    363
    364
    365
    366
    367
    368
    369
    370
    371
    372
    373
    374
    375
    376
    377
    378
    379
    380
    381
    382
    383
    384
    385
    386
    387
    388
    389
    390
    391
    392
    393
    394
    395
    396
    397
    398
    399
    400
    401
    402
    403
    404
    405
    406
    407
    408
    409
    410
    411
    412
    413
    414
    415
    416
    417
    418
    419
    420
    421
    422
    423
    424
    425
    426
    427
    428
    429
    430
    431
    432
    433
    434
    435
    436
    437
    438
    439
    440
    441
    442
    443
    444
    445
    446
    447
    448
    449
    450
    451
    452
    453
    454
    455
    456
    457
    458
    459
    460
    461
    462
    463
    464
    465
    466
    467
    468
    469
    470
    471
    472
    473
    474
    475
    476
    477
    478
    479
    480
    481
    482
    483
    484
    485
    486
    487
    488
    489
    490
    491
    492
    493
    494
    495
    496
    497
    498
    499
    500
    501
    502
    503
    504
    505
    506
    507
    508
    509
    510
    511
    512
    513
    514
    515
    516
    517
    518
    519
    520
    521
    522
    523
    524
    525
    526
    527
    528
    529
    530
    531
    532
    533
    534
    535
    536
    537
    538
    539
    540
    541
    542
    543
    544
    545
    546
    547
    548
    549
    550
    551
    552
    553
    554
    555
    556
    557
    558
    559
    560
    561
    562
    563
    564
    565
    566
    567
    568
    569
    570
    571
    572
    573
    574
    575
    576
    577
    578
    579
    580
    581
    582
    583
    584
    585
    586
    587
    588
    589
    590
    591
    592
    593
    594
    595
    596
    597
    598
    599
    600
    601
    602
    603
    604
    605
    606
    607
    608
    609
    610
    611
    612
    613
    614
    615
    616
    617
    618
    619
    620
    621
    622
    623
    624
    625
    626
    627
    628
    629
    630
    631
    632
    633
    634
    635
    636
    637
    638
    639
    640
    641
    642
    643
    644
    645
    646
    647
    648
    649
    650
    651
    652
    653
    654
    655
    656
    657
    658
    659
    660
    661
    662
    663
    664
    665
    666
    667
    668
    669
    670
    671
    672
    673
    674
    675
    676
    677
    678
    679
    680
    681
    682
    683
    684
    685
    686
    687
    688
    689
    690
    691
    692
    693
    694
    695
    696
    697
    698
    699
    700
   
```

Query Match:	20.2%	Score 24.8;	DB 1;	Length 2064;
Best Local Similarity	57.9%	Pred. No. 7.4;		
Matches 44;	Conservative	0;	Mismatches 32;	Indels 0;
Gaps				0;

QY	8	tgfaaatgtgagatgcttacaacccgatgatgaacaaatgagcaatgcgagggctcact	67
Db	633	tgcttattgttcagcaggaattatcctgcgactggtgaactctaacgctgcacaggggtgtgct	692
QY	68	gactgtgaagtccac	83
Db	693	gactgtgtcagtgscac	708

APPLICANT: NAGASE, KAZUO
 APPLICANT: MATSUI, HIROSHI
 TITLE OF INVENTION: GENE ENCODING TRANSGLUTAMINASE DERIVED
 TITLE OF INVENTION: FROM FISH
 NUMBER OF SEQUENCES: 72
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
 ADDRESSEE: P.C.
 STREET: 1755 Jefferson Davis Highway, Fourth Floor
 City: Arlington
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22202
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/583,799
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/004,729
 FILING DATE: 14-JAN-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Oblon, NO. 5607849man F.
 REGISTRATION NUMBER: 24,618
 REFERENCE/DOCKET NUMBER: 10-599-0
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703)412-3000
 TELEFAX: (703)413-2220
 TELEX: 248855 OPAT UR
 INFORMATION FOR SEQ ID NO: 69:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2148 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA to mRNA
 ORIGINAL SOURCE:
 ORGANISM: Paralichthys olivaceus
 TISSUE TYPE: liver
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 26..2092
 US-08-583-799-69

```

QY      8  tgytaatgtagagatgcttaccacacccctgtagaccctaatgagcgaatgcaaggctgcaact 67
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      661 TGCTATGCTAGCAGCAGGATGATCTACTCCGATGGTGAACCTTAACGCTACACAGGGGTGCT 720

QY      68  gactgtagaagtttcac 83
      ||||| ||| |||
Db      721 GACTGTCACTGGCAC 736

RESULT 13
US-08-583-799-71
; Sequence 71, Application US/08583799
; Patent No. 5607849
; GENERAL INFORMATION:
; APPLICANT: YASUEDA, HISASHI
; APPLICANT: NAKANISHI, KAZUO
; APPLICANT: MOROKI, MASAO
; APPLICANT: MASASE, KAZUO
; APPLICANT: MATSUI, HIROSHI
; TITLE OF INVENTION: GENE ENCODING TRANSGLUTAMINASE DERIVED

```

```

; TITLE OF INVENTION: FROM FISH
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
; STREET: 1755 Jefferson Davis Highway, Fourth Floor
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/583,799
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/004,729
; FILING DATE: 14-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, No. 5607849man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 10-599-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)412-3000
; TELEFAX: (703)413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2148 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; ORIGINAL SOURCE:
; ORGANISM: Paralichthys olivaceus
; TISSUE TYPE: liver
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 26..2092
; US-08-583-799-71.

Query Match 20.2%; Score 24.8; DB 1; Length 2148;
Best Local Similarity 57.9%; Pred. No. 7.6;
Matches 44; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 8 tgaatgtgagatgccttaaaccttgatgactaatgagcggaatgcgaggtcgtcact 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 661 TGCTATGTCAGCAGGATGATCTACCTCGAGGTGTAACGGTGACAGGGGTGTGCT 720
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 68 gactgtgaatttcac 83
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 GACTGTCAGTGCAC 736

RESULT 14
US-09-318-448-5/C
; Sequence 5, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
; APPLICANT: Johnson, William G.
; APPLICANT: Stenroos, Edward S.
; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
; FILE REFERENCE: 601-1-057
; CURRENT APPLICATION NUMBER: US/09/318,448
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.0
```

```

; SEQ ID NO 5
; LENGTH: 7720
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-318-448-5

Query Match 19.8%; Score 24.4; DB 4; Length 7720;
Best Local Similarity 60.6%; Pred. No. 17;
Matches 40; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 2 aggtatgtaaatgtgagatgccttaaacctgatgacctaatggtcgcaatgcgagggc 61
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6888 AGGCTGTGTACCTCGCAGGAAGCATGATACCTGCTGGATCCAGGCCCCCAATTTGGGG 6829
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 62 tgcact 67
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6828 AGCAGCT 6823

RESULT 15
US-08-152-019A-44
; Sequence 44, Application US/08152019A
; Patent No. 5565331
; GENERAL INFORMATION:
; APPLICANT: Tessier-Lavigne, Marc
; APPLICANT: Serafini, Tito
; APPLICANT: Kennedy, Timothy
; APPLICANT: Placzek, Marysia
; APPLICANT: Jessell, Thomas
; APPLICANT: Dodd, Jane
; TITLE OF INVENTION: NEURAL AXON OUTGROWTH MODULATORS
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/152,019A
; FILING DATE: 12-NOV-1993
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman, Richard Aron
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: A-59012/RAO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1839 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-152-019A-44

Query Match 19.7%; Score 24.2; DB 1; Length 1839;
Best Local Similarity 52.5%; Pred. No. 12;
Matches 53; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 1 caggtatgtaaatgtgagatgccttaaacctgtatgactaatggtcgcaatgcgaggg 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 176 CTGTTATTGCTAGTAGATACATGCGGAAACCAACCCAGACAGACAGTATTGTTACTGTGAAG 235
Db 61 ctgcactgactcgtgaagttttacacatgtaaggtgaacttat 101
Db 236 AGGGTCCGGATGGAATTATTCGCTAGCAGATGTGACACTTGT 276

```

US-08-482-677-1          RESULT 16
US-08-482-677-1          Sequence 1, Application US/08482677
Patent No. 6017714
GENERAL INFORMATION:
APPLICANT: Tessier-Lavigne, Marc
APPLICANT: Serafini, Tito
APPLICANT: Kennedy, Timothy
APPLICANT: Placzek, Marysia
APPLICANT: Jessel, Thomas
APPLICANT: Dodd, Jane
TITLE OF INVENTION: Netlins
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESSES:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.3
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/482,677
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UC93-300-4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1839 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-482-677-1

```

Query Match	19.7%	Score 24.2	DB 3	Length 1839
Best Local Similarity	52.5%	Pred. No. 12		
Matches 53; Conservative	0	Mismatches 48	Indels 0	Gaps 0

QY 61 ctgcactgactgtgaagttccacatgtaagtgatacttat 101
| | | | | | | | | | | | | | | | | | | | |
Db 236 AGGGTCCGATGGAAATATCCGTGAGCAATGTGACACTTGT 276

RESULT 17
US-08-727-034-5
; Sequence 5, Application US/08727034
; Patent No. 5655872
; GENERAL INFORMATION:
; APPLICANT: SATTO, YASHUSHI
; APPLICANT: IMASAKI, AKIO

```

APPLICANT : ARAI, KOICHI
APPLICANT : YAMAZAKI, HIROYUKI
TITLE OF INVENTION: NOVEL LDL RECEPTOR ANALOG PROTEIN AND
TITLE OF INVENTION: THE GENE CODING THEREFOR
NUMBER OF SEQUENCES:      8
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/727,034
FILING DATE: 08-OCT-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 261440/1995
FILING DATE: 09-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 102451/1996
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 80-079-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 6642 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
IS-08-727-034-5

```

Query Match	19.7%;	Score 24.2;	DB 1;	Length 6642;
Best Local Similarity	52.5%;	Pred. No. 19;		
Matches	53;	Conservative	0;	Mismatches 48; Indels 0; Gaps 0;

Oy	63	gcactgactgtaagttccacatgtaagtgatactatct	103
Db	5248	ccaccagatatccacattgacacgctatggtgaaaattatct	5288

RESULT 18
 US-08-727-034-6
 ? Sequence 6, Application US/08727034
 ? Patent No. 5665872
 ?
 ? GENERAL INFORMATION:
 ?
 ? APPLICANT: SATTO, YASHUSHI
 ? APPLICANT: IMASAKI, AKIO
 ?
 ? APPLICANT: ARAI, KOICHI
 ?
 ? APPLICANT: YAMAZAKI, HIROYUKI
 ?
 ? TITLE OF INVENTION: NOVEL LDL RECEPTOR ANALOG PROTEIN AND
 ?
 ? TITLE OF INVENTION: THE GENE CODING THEREFOR
 ?
 ? NUMBER OF SEQUENCES: 8
 ?
 ? CORRESPONDENCE ADDRESS:
 ?
 ? ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
 ?
 ? ADDRESSEE: P.C.

```
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/727,034
FILING DATE: 08-OCT-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 261440/1995
FILING DATE: 09-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 102451/1996
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 80-079-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-413-3000
TELEFAX: 703-413-2220
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 6843 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA to mRNA
FEATURE:
NAME/KEY: CDS
LOCATION: 81..6725
OTHER INFORMATION: /note= "Identification Method: S"
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 81..164
OTHER INFORMATION: /note= "Identification Method: S"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 165..6722
OTHER INFORMATION: /function= "Nucleotides 165-6722
encode the mature peptide"
OTHER INFORMATION: /note= "Identification Method: S"
US-08-727-034-6

Query Match          19.7%; Score 24.2; DB 1; Length 6843;
Best Local Similarity 52.5%; Pred. No. 19;
Matches 53; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 3 ggatgtgaatgagtcgacccctacaccccgatgacctaagggcgcaatggcgagggct 62
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5268 GGAATGGAAGTGGAGCGCATCTTAATCCATTACCAACCAAGAAAGATGCCCA 5327

QY 63 gcactgactgtaagtttcacatgttaaggtgactatct 103
    ||| | | | | | | | | | | | | | | | | | | | | |
DB 5328 CCACGAGATATCCACATTCAGACGCTATGCTGAATAATTATCT 5368

RESULT 19
US-08-253-785-2/c
; Sequence 2, Application US/08253785
; Patent No. 5633363
; GENERAL INFORMATION:
; APPLICANT: Colbert, James T.
; APPLICANT: Held, Bruce M.
; APPLICANT: Wurtel, Eve S.
; APPLICANT: Dietrich, Paul S.
```

```
TITLE OF INVENTION: ROOT PREFERENTIAL PROMOTER
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Sandoz Agro, Inc.
STREET: 975 California Avenue
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/253,785
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Marcus-Wyner, Lynn
REGISTRATION NUMBER: 34,869
REFERENCE/DOCKET NUMBER: 135-1089
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/354-3588
TELEFAX: 415/857-1125
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2329 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 51..1850
OTHER INFORMATION: /codon_start= 51
US-08-253-785-2

Query Match          19.2%; Score 23.6; DB 1; Length 2329;
Best Local Similarity 56.4%; Pred. No. 21;
Matches 44; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 10 taatgtgagatgccttacacccctgacctaagcgcgcaatggcgaggtcgactga 69
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1001 TAAATCTTCATGCTTGGAGACCATATGCAATTGTGGCTTATTGATCATACACACC 942

QY 70 ctgtaagttcacatgt 87
    ||| | | | | | | | | | | | | | | | | | | | | |
DB 941 ATGGACCCCATATGATGT 924

RESULT 20
US-08-975-316-53/c
; Sequence 53, Application US/08975316
; Patent No. 5952486
; GENERAL INFORMATION:
; APPLICANT: BLOKSBERG, Leonard N., HAYUKKALA, Ilkka
; APPLICANT: and GRIERSON, Alastair W.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Offices of Ann W. Speckman
; STREET: 2601 Elliott Avenue, Suite 4185
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
```


	Query Match	18.7%	Score 23	DB 4	Length 1959
	Best Local Similarity	74.4%	Pred. No. 33		
	Matches 29	Conservative 0	Mismatches 10	Indels 0	Gaps 0
QY	10 taatgtgagatgccttaacaacctgatacctaatagc	48			
db	1005 tcaatatgataagaccttaacaacctgggttaacaatgc	1043			

Query	Best Match	Query Similarity	Score	DB	Length	Z-score
1322	ATTTGGAGACATCGGGGCTCTCGCTCTCTAAGGTAGGACATTTTAAAGGATATTATTC	55.7%	18.7	DB	1381	2059
103	ttcaatttatttcacacaa	0	0	0	0	0

Db 1382 CTCGAGATTCCTATACCTA 1400

RESULT 26

US-08-369-796-7/C
 ; Sequence 7, Application US/08369796
 ; Patent No. 5716622
 ; GENERAL INFORMATION:
 ; APPLICANT: James E. Darnell, Jr.
 ; APPLICANT: Zilong Wen
 ; APPLICANT: Curt M. Horvath
 ; APPLICANT: Zhong Zhong
 ; TITLE OF INVENTION: FUNCTIONALLY ACTIVE REGIONS OF SIGNAL
 ; TITLE OF INVENTION: TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION (STAT) PROTEINS
 ; NUMBER OF SEQUENCES: 39
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Klauber & Jackson
 ; STREET: 411 Hackensack Avenue
 ; CITY: Hackensack
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07601
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/369,796
 ; FILING DATE: 06-JAN-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Jackson Esq., David A.
 ; REGISTRATION NUMBER: 26,742
 ; REFERENCE/DOCKET NUMBER: 600-1-116
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 201 487-5800
 ; TELEFAX: 201 343-1684
 ; TELEX: 133521
 ; INFORMATION FOR SEQ ID NO: 7:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2277 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: both
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: CDNA
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Mouse
 ; IMMEDIATE SOURCE:
 ; CLONE: Murine Stat91
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 5..2251
 ; US-08-369-796-7

Query Match 18.7%; Score 23; DB 1; Length 2277;
 Best Local Similarity 68.1%; Pred. No. 35;
 Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 70 ctgtaagtttcacatgtaagtgatacttacttcaatttca 116
 ||| | |||| | |||| | |||| | |||| | |||| | |||| |
 Db 538 CTGAGAGGTTTGCAATTAAGTCATATTCATCTGTATCTCTCA 492

RESULT 27
 US-08-852-091-7/C
 ; Sequence 7, Application US/08852091
 ; Patent No. 5883228
 ; GENERAL INFORMATION:

APPLICANT: James E. Darnell, Jr.
 APPLICANT: Zilong Wen
 APPLICANT: Curt M. Horvath
 APPLICANT: Zhong Zhong
 TITLE OF INVENTION: FUNCTIONALLY ACTIVE REGIONS OF SIGNAL
 TITLE OF INVENTION: TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION (STAT) PROTEINS
 NUMBER OF SEQUENCES: 39
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Klauber & Jackson
 STREET: 411 Hackensack Avenue
 CITY: Hackensack
 STATE: New Jersey
 COUNTRY: USA
 ZIP: 07601
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/852,091
 FILING DATE: 06-MAY-1997
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/369,796
 FILING DATE: 06-JAN-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Jackson Esq., David A.
 REGISTRATION NUMBER: 26,742
 REFERENCE/DOCKET NUMBER: 600-1-116
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201 487-5800
 TELEFAX: 201 343-1684
 TELEX: 133521
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2277 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: both
 TOPOLOGY: unknown
 MOLECULE TYPE: CDNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Mouse
 IMMEDIATE SOURCE:
 CLONE: Murine Stat91
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 5..2251
 ; US-08-852-091-7

Query Match 18.7%; Score 23; DB 2; Length 2277;
 Best Local Similarity 68.1%; Pred. No. 35;
 Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 70 ctgtaagtttcacatgtaagtgatacttacttcaatttca 116
 ||| | |||| | |||| | |||| | |||| | |||| | |||| |
 Db 538 CTGAGAGGTTTGCAATTAAGTCATATTCATCTGTATCTCTCA 492

RESULT 28
 US-08-820-754-7/C
 ; Sequence 7, Application US/08820754
 ; Patent No. 5976835
 ; GENERAL INFORMATION:
 ; APPLICANT: Darnell Jr., James E.
 ; APPLICANT: Schindler, Christian W.
 ; APPLICANT: Fu, Xian-Yuan
 ; APPLICANT: Wen, Zilong
 ; APPLICANT: Zhong, Zhong
 ; TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN


```

TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauder & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/820,754
FILING DATE: 19-MAR-1997
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,185
FILING DATE: 11-MAR-1994
APPLICATION NUMBER: US 07/980,498
FILING DATE: 23-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/854,296
FILING DATE: 19-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO US93/02569
FILING DATE: 19-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/126,588
FILING DATE: 24-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-073 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 467-5800
TELEFAX: 201 343-1684
TELEX: 133521

INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mouse
IMMEDIATE SOURCE:
CLONE: Murine Stat91
FEATURE:
NAME/KEY: CDS
LOCATION: 5..2251
US-08-820-754-7

Query Match 18.7%; Score 23; DB 2; Length 2277;
Best Local Similarity 68.1%; Pred. NO. 35;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Oy 70 ctgtaagtttcacatgtaagtgatctatcttcacattatccca 116
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 538 ctgacaggtttgcatttaaaagtcattatcttctgtattcttcta 492

RESULT 29
US-08-956-652-7/c
Sequence 7, Application US/08956652
Patent No. 6013475
GENERAL INFORMATION:

```

```

APPLICANT: Darnell Jr., James E.
APPLICANT: Schindler, Christian W.
APPLICANT: Fu, Xian-tuan
APPLICANT: Wen, Zilong
APPLICANT: Zhong, Zhong
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
TITLE OF INVENTION: SEQUENCES AND METHODS OF USE THEREOF
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey.
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,652
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/212,185
FILING DATE: 11-MAR-1994
APPLICATION NUMBER: US 07/980,498
FILING DATE: 23-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/854,296
FILING DATE: 19-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO US93/02569
FILING DATE: 19-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/126,588
FILING DATE: 24-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-073 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE: Mouse
ORGANISM: Mouse
IMMEDIATE SOURCE:
CLONE: Murine Stat9l
FEATURE:
NAME/KEY: CDS
LOCATION: 5..2251
US-08-956-652-7
Query Match 18.7%; Score 23; DB 3; Length 2277;
Best Local Similarity 68.1%; Pred. NO. 35;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
0y 70 ctgtaagcttcacatgtaaggtagtacttatcctaattatcca 116
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
538 ctgagagcttttgcaattttaagctcatatttcattcttgtaatcttcta 492

```

RESULT 30
US-08-956-869-7/c
Sequence 7, Application US/08956869
Patent No. 6030808
GENERAL INFORMATION:
APPLICANT: Darnell Jr., James E.
APPLICANT: Schindler, Christian W.
APPLICANT: Fu, Xian-Yuan
APPLICANT: Wen, Zilong
APPLICANT: Zhong, Zhong
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/956,869
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/212,185
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/854,296
FILING DATE: 19-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO US93/02569
FILING DATE: 19-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/126,588
FILING DATE: 24-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-073 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mouse
IMMEDIATE SOURCE:
CLONE: Murine Stat91
FEATURE:
NAME/KEY: CDS
LOCATION: 5..2251
US-08-956-869-7

Query Match 18.7%; Score 23; DB 3; Length 2277;
Best Local Similarity 68.1%; Pred. No. 35;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 70 ctgtaagltcacatgtaagtgatcattatcttaatttca 116
Db 538 CTGAGAGTTTGCAATTAAAGTCATATTCATCTTGTAATCTTCTA 492
RESULT 31
US-08-948-547-7/c
Sequence 7, Application US/08948547
Patent No. 6124118
GENERAL INFORMATION:
APPLICANT: Darnell Jr., James E.
APPLICANT: Schindler, Christian W.
APPLICANT: Fu, Xian-Yuan
APPLICANT: Wen, Zilong
APPLICANT: Zhong, Zhong
TITLE OF INVENTION: RECEPTOR RECOGNITION FACTORS, PROTEIN
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/948,547
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/212,185
FILING DATE: 11-MAR-1994
APPLICATION NUMBER: US 07/980,498
FILING DATE: 23-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/854,296
FILING DATE: 19-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO US93/02569
FILING DATE: 19-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/126,588
FILING DATE: 24-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-073 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Mouse
IMMEDIATE SOURCE:
CLONE: Murine Stat91
FEATURE:
NAME/KEY: CDS
LOCATION: 5..2251
US-08-948-547-7

Query Match 18.7%; Score 23; DB 3; Length 2277;
Best Local Similarity 68.1%; Pred. No. 35;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 70 ctgtgaatttcacatgtaagtgatactatcttcattatcca 116
111 11111 11111 11111 11111 11111 11111 11111 11111 11111
Db 538 CTGAGAGTTTTCATTTAAAGTCATATTCATCTTGTAAATTCCTCTCA 492

RESULT 32

PCT-US95-17025-7/c

; Sequence 7, Application PC/TUS9517025

; GENERAL INFORMATION:

; APPLICANT: James E. Darnell, Jr.

; APPLICANT: Zilong Wen

; APPLICANT: Curt M. Horvath

; APPLICANT: Zhong Zhong

; TITLE OF INVENTION: FUNCTIONALLY ACTIVE REGIONS OF SIGNAL

; TITLE OF INVENTION: TRANSDUCER AND ACTIVATOR OF TRANSCRIPTION (STAT) PROTEINS

; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Klauber & Jackson

; STREET: 411 Hackensack Avenue

; CITY: Hackensack

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US95/17025

; FILING DATE: 28-DEC-1995

; CLASSIFICATION:

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: 08/369,796

; FILING DATE: 06-JAN-1995

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Jackson Esq., David A.

; REGISTRATION NUMBER: 26,742

; REFERENCE/DOCKET NUMBER: 600-1-116

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201 487-5800

; TELEFAX: 201 343-1684

; TELEX: 133521

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2277 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: both

; TOPOLOGY: unknown

; MOLECULE TYPE: cDNA

; HYPOTHEICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Mouse

; IMMEDIATE SOURCE:

; CLONE: Murine Stat91

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 5..2251

; PCT-US95-17025-7

Query Match 18.7%; Score 23; DB 5; Length 2277;
Best Local Similarity 68.1%; Pred. No. 35;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 70 ctgtgaatttcacatgtaagtgatactatcttcattatcca 116
111 11111 11111 11111 11111 11111 11111 11111 11111 11111
Db 538 CTGAGAGTTTTCATTTAAAGTCATATTCATCTTGTAAATTCCTCTCA 492

RESULT 33

US-09-318-448-30

; Sequence 30, Application US/09318448

; Patent No. 6210950

; GENERAL INFORMATION:

; APPLICANT: Johnson, William G.

; APPLICANT: Stenroos, Edward S.

; TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING

; TITLE OF INVENTION: DEVELOPMENTAL DISORDERS

; FILE REFERENCE: 601-1-057

; CURRENT APPLICATION NUMBER: US/09/318,448

; CURRENT FILING DATE: 1999-05-25

; NUMBER OF SEQ ID NOS: 46

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 30

; LENGTH: 3464

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-09-318-448-30

Query Match 18.7%; Score 23; DB 4; Length 3464;
Best Local Similarity 54.0%; Pred. No. 41;
Matches 47; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 1 cagatattgtaaatgtgagatgccttaaacacctgagacctaagcgcaatgaggg 60
111 11111 11111 11111 11111 11111 11111 11111 11111 11111
Db 331 cagacactgttaacagtgagctgcacatccacacacatcatgtgacacttgcagtc 390

QY 61 ctgcactgactggttaagttcacatgt 87
11 11 1111111111
Db 391 ctgcactgtgcgacttcacatgt 417

RESULT 34

US-09-056-105-27

; Sequence 27, Application US/09056105

; Patent No. 6287569

; GENERAL INFORMATION:

; APPLICANT: KIPPS, THOMAS J.

; APPLICANT: Wu, YUNQI

; TITLE OF INVENTION: VACCINES WITH ENHANCED INTRACELLULAR

; FILE REFERENCE: 233/221

; CURRENT APPLICATION NUMBER: US/09/056,105

; CURRENT FILING DATE: 1998-04-06

; EARLIER APPLICATION NUMBER: 60/043,467

; EARLIER FILING DATE: 1997-04-10

; NUMBER OF SEQ ID NOS: 35

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 27

; LENGTH: 4233

; TYPE: DNA

; ORGANISM: Homo sapiens

; US-09-056-105-27

Query Match 18.7%; Score 23; DB 4; Length 4233;
Best Local Similarity 60.3%; Pred. No. 45;
Matches 38; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 7 ttgttaatgtgagatgcttaaacacctgagacctaagcgcaatgagggctgcaac 66
11111111 11111111 11111111 11111111 11111111 11111111 11111111 11111111
Db 2910 ttgaaatgactgtactatataccctctctacacttaggcctcttcttgtygaacac 2969
QY 67 tga 69
111
Db 2970 tga 2972

```
RESULT 35
US-07-792-600-1/c
; Sequence 1, Application US/07792600
; Patent No. 6008045
; GENERAL INFORMATION:
; APPLICANT: COPELAND, WILLIAM C.
; APPLICANT: WANG, TERESA S.-F.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Peter G. Carroll
; STREET: 220 Montgomery Street, Suite 710
; CITY: San Francisco
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/792,600
; FILING DATE: 19911115
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4440 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-07-792-600-1

Query Match      18.7%; Score 23; DB 3; Length 4440;
Best Local Similarity 68.1%; Pred. No. 45;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 3 ggtattgtaaatgtgagatgccttaacaacccctgatactaatgagcg 49
    ||| |||| | | | | | | | | | | | | | | | | | | | | |
DB 408 GGTATTGTAACACTGCAGCCTTCTTACATTCCCTGTGCTTTATTGCG 362

RESULT 36
US-09-157-021-1/c
; Sequence 1, Application US/09157021A
; Patent No. 6100023
; GENERAL INFORMATION:
; APPLICANT: Copeland, William C.
; APPLICANT: Wang, Teresa S. F.
; TITLE OF INVENTION: Drug Design Assay
; FILE REFERENCE: STDU-03484
; CURRENT APPLICATION NUMBER: US/09/157,021A
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: 07/792,600
; EARLIER FILING DATE: 1991-11-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4440
; TYPE: DNA
; ORGANISM: Homo sapiens
;
US-09-157-021-1
```

```
Query Match      18.7%; Score 23; DB 3; Length 4440;
Best Local Similarity 68.1%; Pred. No. 45;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 3 ggtattgtaaatgtgagatgccttaacaacccctgatactaatgagcg 49
    ||| |||| | | | | | | | | | | | | | | | | | | | | |
DB 408 GGTATTGTAACACTGCAGCCTTCTTACATTCCCTGTGCTTTATTGCG 362

RESULT 37
US-09-156-842-1/c
; Sequence 1, Application US/09156842A
; Patent No. 6103473
; GENERAL INFORMATION:
; APPLICANT: Copeland, William C.
; APPLICANT: Wang, Teresa S. F.
; TITLE OF INVENTION: Drug Screening
; FILE REFERENCE: STDU-03485
; CURRENT APPLICATION NUMBER: US/09/156,842A
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: 07/792,600
; EARLIER FILING DATE: 1991-11-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 4440
; TYPE: DNA
; ORGANISM: Homo sapiens
;
US-09-156-842-1

Query Match      18.7%; Score 23; DB 3; Length 4440;
Best Local Similarity 68.1%; Pred. No. 45;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 3 ggtattgtaaatgtgagatgccttaacaacccctgatactaatgagcg 49
    ||| |||| | | | | | | | | | | | | | | | | | | | | |
DB 408 GGTATTGTAACACTGCAGCCTTCTTACATTCCCTGTGCTTTATTGCG 362

RESULT 38
US-09-157-021-35/c
; Sequence 35, Application US/09157021A
; Patent No. 6100023
; GENERAL INFORMATION:
; APPLICANT: Copeland, William C.
; APPLICANT: Wang, Teresa S. F.
; TITLE OF INVENTION: Drug Design Assay
; FILE REFERENCE: STDU-03484
; CURRENT APPLICATION NUMBER: US/09/157,021A
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: 07/792,600
; EARLIER FILING DATE: 1991-11-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 35
; LENGTH: 5433
; TYPE: DNA
; ORGANISM: Homo sapiens
;
US-09-157-021-35

Query Match      18.7%; Score 23; DB 3; Length 5433;
Best Local Similarity 68.1%; Pred. No. 49;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 3 ggtattgtaaatgtgagatgccttaacaacccctgatactaatgagcg 49
    ||| |||| | | | | | | | | | | | | | | | | | | | | |
DB 408 GGTATTGTAACACTGCAGCCTTCTTACATTCCCTGTGCTTTATTGCG 362
```

RESULT 39
US-09-156-842-35/C
; Sequence 35, Application US/09156842A
; Patent No. 6103473
; GENERAL INFORMATION:
; APPLICANT: Copeland, William C.
; APPLICANT: Wang, Teresa S. F.
; TITLE OF INVENTION: Drug Screening
; FILE REFERENCE: STDU-03485
; CURRENT APPLICATION NUMBER: US/09/156,842A
; CURRENT FILING DATE: 1998-09-18
; EARLIER APPLICATION NUMBER: 07/792,600
; EARLIER FILING DATE: 1991-11-15
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 35
; LENGTH: 5433
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-156-842-35

Query Match 18.7%; Score 23; DB 3; Length 5433;
Best Local Similarity 68.1%; Pred. No. 49;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 3 ggtatgttaattgtgagcttcaacacctgatgacctaatggcg 49
DB 408 GGTGTTGTCTACTGCGAGCTCTTACATCTCTTGCTTATTATGCG 362

RESULT 40
US-08-483-503A-2
; Sequence 2, Application US/08483503A
; Patent No. 601/880
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, Stephen
; APPLICANT: Wahl, Sharon M.
; APPLICANT: Thompson, Robert C.
; APPLICANT: Drilpe, David J.
; TITLE OF INVENTION: INHIBITION OF RETROVIRUS INFECTION
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,503A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/209,040
; FILING DATE: 09-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,369
; FILING DATE: 09-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Barker, M. Paul
; REGISTRATION NUMBER: 32,013
; REFERENCE/DOCKET NUMBER: 04189,0084-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
; LENGTH: 180 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-483-503A-2

Query Match 18.5%; Score 22.8; DB 3; Length 180;
Best Local Similarity 50.9%; Pred. No. 16;
Matches 54; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

OY 5 tatgttaattgtgagcttcaacacctgatgacctaatggcgcaatggcgctgc 64
DB 61 TATGCCCAATGTTGATGCTTACACCCCAATTTCTGTGAGATGATGCCCAAGTCAAG 120
OY 65 actgactgttaattgtgagcttcaacacctgatgacctaatggcgcaatggcgctgc 110
DB 121 CGTGACTTGAAGTGTGATGCGCATGTGTGGAAATCTGCGCTT 166

RESULT 41
US-08-896-605A-7/C
; Sequence 7, Application US/08896605A
; Patent No. 5879942
; GENERAL INFORMATION:
; APPLICANT: TANIMOTO, Tadao
; APPLICANT: KURIMOTO, Masashi
; TITLE OF INVENTION: PROCESSING ENZYME FOR POLYPEPTIDE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEWMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/896,605A
; FILING DATE: 18 July 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 207,691/1996
; FILING DATE: 19-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 156,062/1997
; FILING DATE: 30-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TANIMOTO=2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 579 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; FEATURE:
; NAME/KEY: leader peptide
; LOCATION: 1..108
; IDENTIFICATION METHOD: S
; NAME/KEY: mat peptide
; LOCATION: 109..579
; IDENTIFICATION METHOD: S


```

; APPLICANT: KORRKO, JARMO
; APPLICANT: ALA-KORRKO, LEENA, et al.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
; TITLE OF INVENTION: ALTERED TYPE I OR TYPE IX COLLAGEN GENE SEQUENCES
; NUMBER OF SEQUENCES: 666
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PANITCH SCHWARZE JACOBS & NADEL, P.C.
; STREET: ONE COMMERCE SQUARE, 2005 MARKET STREET, 22ND
; STREET: FLR
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-7086
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/943,711
; FILING DATE: 03-OCT-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/212,322
; FILING DATE: 14-MAR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/803,628
; FILING DATE: 03-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: DOYLE LEARY Ph.D., KATHRYN
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: 9598-27
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-965-1284
; TELEFAX: 215-567-2991
; TELEX: 831-494
; INFORMATION FOR SEQ ID NO: 81:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 911 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-943-731-81

Query Match 18.5%; Score 22.8; DB 4; Length 911;
Best Local Similarity 66.0%; Pred. No. 30;
Matches 33; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

QY 64 cactgactgtaagttaacatgtaagtgatctatctcaattatt 113
DB 774 CACTAGCTGGGAAGTGTGAAGAAGATTTCTCTGATATTTT 725

RESULT 45
US-08-884-324-2/c
; Sequence 2, Application US/08884324
; Patent No. 6060283
; GENERAL INFORMATION:
; APPLICANT: Takamori OKURA
; APPLICANT: Kakuji TORIGOE
; APPLICANT: Masahi KURIMOTO
; TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEWMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/884,324
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 185,305/96
; FILING DATE: 27-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: OKURA-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1120 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; MOLECULE TYPE: cDNA to mRNA
; TOPOLOGY: linear
; HYPOHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: human
; TISSUE TYPE: liver
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..177
; IDENTIFICATION METHOD: E
; NAME/KEY: leader peptide
; LOCATION: 178..285
; IDENTIFICATION METHOD: S
; NAME/KEY: mat peptide
; LOCATION: 286..756
; IDENTIFICATION METHOD: S
; NAME/KEY: 3'UTR
; LOCATION: 757..1120
; IDENTIFICATION METHOD: E
; US-08-884-324-2

Query Match 18.5%; Score 22.8; DB 3; Length 1120;
Best Local Similarity 50.0%; Pred. No. 32;
Matches 57; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 3 ggtattgtaaatgtgagatgccttaacaacccgatgacctaattgacgaaggct 62
DB 337 GGTCATTCAAATTTCTTATGACGATTAATTAGATTCAGCTGCCAAGTAATCTGAT 278

QY 63 gcaactgactgtaagttaacatgtaagtgatctatctcaattatt 116
DB 277 CCAGGTTTCATCATCTTCAGCTATTAAGTAAGCGATATGTCATAAATTTCA 224
```

Search completed: March 19, 2002, 09:58:41
Job time: 1615 sec

OM of: US-09-684-016-48411 to: PIR_68:* out_format : pfs

Date: Mar 19, 2002 9:32 AM

About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+np2 model -DEV=x1h
-O=/cgn2.1/USPO.spool/US09684016/unat_19032002_084440_20919/app_query.fasta.1.182
-DB=PIR_68 -GFF=fastan -SUFFIX=std.rpr -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPL=0.000 -LOOEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-DELPO=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELPO=6.000 -DELEXT=7.000 -START=1 -MATRIX=Blustum62
-THRS=human40.cdi -LIST=45 -DOCALLGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NOR=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09684016.ecgn1.1.56 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-684-016-48411

Query Length: 123

Database: PIR_68:*

Database sequences: 219241

Database length: 76174552

Search time (sec): 37.240000

score_list:

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
PIR2:C85253	+	138.00	341.52	201	receptor like protein (partial)
PIR2:T49121	+	138.00	341.52	201	receptor like protein - Arabid
PIR2:S44281	+	127.00	312.10	227	ES43 protein - barley
PIR2:T06070	+	124.00	304.12	258	ES43 protein homolog F19H22.200
PIR2:H85462	+	124.00	304.12	258	ES43 like protein [imported]
PIR2:S69047	+	66.50	155.12	353	hypothetical protein YP138c -
PIR2:T40419	+	65.50	153.03	329	mi02 protein - fission yeast (S
PIR2:T41449	+	62.50	143.69	122	probable phd finger transcript
PIR2:T21433	+	62.00	141.99	142	hypothetical protein F26H11.3a
PIR2:T21435	+	62.00	141.99	142	hypothetical protein F26H11.3c
PIR2:T21430	+	62.00	141.19	510	hypothetical protein F26H11.3b
PIR2:T03455	+	60.00	121.04	191	ALR protein - human
PIR2:T03454	+	60.00	120.64	187	ALR protein - human
PIR2:T25101	+	59.50	136.15	3.27	hypothetical protein T22C1.1 -
PIR2:T08905	+	59.50	135.74	3.24	hypothetical protein T32A16.30
PIR2:H81911	+	57.00	139.99	4.33	probable molybdopterin-guanine
PIR2:S48729	+	56.50	125.86	8.22	probable transcription regulat
PIR2:H72647	+	56.00	125.44	9.88	glucokinase regulator - African
PIR2:T42156	+	55.00	119.73	8.67	5-oxoprolinase (ATP-hydrolyz
PIR2:T32671	+	54.50	114.24	13.64	hypothetical protein Y47H9C.2 -
PIR2:T32633	+	53.00	125.09	31.21	hypothetical protein F1561.1 -
PIR2:A31788	+	53.00	124.42	30.74	polypeptide (PC 3.4.99.-)
PIR2:B00542	+	52.00	122.04	42.83	hypothetical protein modA [impc
PIR2:B86073	+	52.00	122.04	42.83	hypothetical protein modA [impc
PIR2:S40803	+	52.00	120.63	41.47	ABC transporter, ATP-binding pr
PIR2:H70109	+	52.00	113.35	35.10	hypothetical protein SPC645.13
PIR2:T41530	+	52.00	112.59	34.49	probable sucrose synthase (EC 2
PIR2:T06497	+	51.50	117.35	46.66	exogastrolin-inducing protein (E
PIR2:A40084	+	51.50	117.35	46.66	exogastrolin-inducing protein (E
PIR2:T38168	+	51.50	110.17	36.94	hypothetical protein SPAC22E12.
PIR2:D09660	+	51.00	124.21	66.22	hypothetical protein Y37A1B.4 -
PIR2:T26611	+	51.00	119.12	58.94	anaerobic dimethylsulfoxide red
PIR2:T64828	+	51.00	119.12	58.94	hypothetical protein dmsB [impc
PIR2:T885615	+	51.00	117.96	57.37	hypothetical protein F32D8.3 -
PIR2:T04224	+	51.00	114.74	53.34	adenylsuccinate lyase (EC 4.3.1
PIR2:A35291	+	51.00	113.79	52.16	ABC transporter, ATP-binding pr
PIR2:F71844	+	51.00	112.26	50.37	ABC transporter, ATP-binding pr

PIR2:H86457	+	51.00	111.02	48.96	697	hypothetical protein AAG2160
PIR2:T15670	+	50.50	105.46	52.29	1332	hypothetical protein C27H5
PIR2:S22412	+	50.00	117.42	83.42	180	signal peptidase (EC 3.4.99
PIR2:S61299	-	50.00	113.29	75.89	336	lipopolysaccharide heptosylt
PIR2:B81669	-	50.00	113.29	75.89	336	lipopolysaccharide heptosylt

seq_name: PIR2:C85253

seq_documentation block:

receptor like protein (partial) [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: C85253
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp
Nature 402, 769-777, 1999
A:title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:reference number: A85001; MIM:20083488
A:Accession: C85253
A:Status: preliminary
A:molecule type: DNA
A:Residues: 1-201 <STO>
A:Cross-references: GB:NC_001268; NID:97269059; PID:CA879169.1; GSPDB:GN00140
C:Genetics:
A:gene: AT4G22140
A:Map position: 4

alignment_scores:

Quality:	138.00	Length:	23
Ratio:	6.273	Gaps:	0
Percent Similarity:	95.652	Percent Identity:	91.304

alignment block:

US-09-684-016-48411 x C85253 ..
Align seq 1/1 to: C85253 from: 1 to: 201

```
5 TATGTAAAGTGGATGAGCTTACACCGTACGACCTAATGGCGCATG 54
|||||
115 TTYCTSYSCYSGIUMETPTOTYRASHNPASPSPLEUETVALGINCY 131
|||||
131 SGUUGLYCYSLYSASPTTP 137
```

seq_name: PIR2:T49121

seq_documentation block:

receptor like protein - Arabidopsis thaliana (fragment)
N:Alternate names: protein AT4G22140
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T49121
R:Bevan, M.; Weiler, H.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K
submitted to the Protein Sequence Database, May 2000
A:reference number: Z25016
A:Accession: T49121
A:Status: preliminary
A:molecule type: DNA
A:Residues: 1-201 <BEV>
A:Cross-references: EMBL:AL022140; GSPDB:GN00062; ATSP:AT4G22140
C:Genetics:
A:gene: ATSP:AT4G22140
A:Map position: 4
A>Note: Intron positions not resolved

alignment_scores:

Quality:	138.00	Length:	23
Ratio:	6.273	Gaps:	0
Percent Similarity:	95.652	Percent Identity:	91.304

alignment_block:

US-09-684-016-48411 x T49121 ..

Align seg 1/1 to: T49121 from: 1 to: 201

5 TATTGTAATGTGAGATGCTTACACCTGATGACCTAATGCGCAATG 54
|||||
115 TYRCYLSYSCYSGlUmetProTyrAsnProAspAspLeuMetValGlnCy 131
|||||
55 CGAGGGCTGCACCTGACTGG 73
|||||
131 sglUglUCySserAspTTP 137

seq_name: pir2:S44281

seq_documentation_block:

ES43 protein - barley

C:Species: Hordeum vulgare (barley)

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 24-Mar-1999

C:Accession: S44281

R:Speulman, E.; Salami, F.
submitted to the EMBL Data Library, February 1994

A:Reference number: S44281

A:Accession: S44281

A:Molecule type: mRNA

A:Residues: 1-227 <SPE>

A:Cross-references: EMBL:X77575; NID:9482918; PID:q1345528

C:Genetics:

A:Gene: ES43

alignment_scores:

Quality:	127.00	Length:	23
Ratio:	6.048	Gaps:	0
Percent Similarity:	91.304	Percent Identity:	82.609

alignment_block:

US-09-684-016-48411 x S44281 ..

Align seg 1/1 to: S44281 from: 1 to: 227

5 TATTGTAATGTGAGATGCTTACACCTGATGACCTAATGCGCAATG 54
:::|||||
147 PheCYLSYSCYSGlUmetProTyrAsnProAspAspLeuMetValGlnCy 163
|||||
55 CGAGGGCTGCACCTGACTGG 73
|||||
163 sglUglUCySserAspTTP 169

seq_name: pir2:T06070

seq_documentation_block:

ES43 protein homolog F19H22.200 - Arabidopsis thaliana (fragment)

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 22-Oct-1999

C:Accession: T06070

R:Bevan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X
submitted to the Protein Sequence Database, March 1999

A:Reference number: Z15184

A:Accession: T06070

A:Molecule type: DNA

A:Residues: 1-258 <BEV>

A:Cross-references: EMBL:AL035679; GSPDB:GN00062; ATSP:F19H22.200

A:Experimental source: cultivar Columbia; BAC clone F19H22

C:Genetics:

A:Gene: ATSP:F19H22.200

A:Map position: 4

A:Introns: 25/1; 140/2; 163/2

alignment_scores:

Quality:	124.00	Length:	23
Ratio:	5.636	Gaps:	0

Percent Similarity: 95.652 Percent Identity: 78.261

alignment_block:

US-09-684-016-48411 x T06070 ..

Align seg 1/1 to: T06070 from: 1 to: 258

5 TATTGTAATGTGAGATGCTTACACCTGATGACCTAATGCGCAATG 54
:::|||||
141 PheCYLSYSCYSGlUmetProTyrAsnProAspAspLeuMetValGlnCy 157
|||||
55 CGAGGGCTGCACCTGACTGG 73
|||||
157 sglUglUCySserGluTTP 163

seq_name: pir2:H85462

seq_documentation_block:

ES43 like protein (imported) - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001

C:Accession: H85462

R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp
Mature 402, 769-777, 1999

A>Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488

A:Accession: H85462

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-258 <STO>

A:Cross-references: GB:NC_001268; NID:97270893; PIDN:CAB80573.1; GSPDB:GN00140

C:Genetics:

A:Gene: AT4g39100

A:Map position: 4

alignment_scores:

Quality:	124.00	Length:	23
Ratio:	5.636	Gaps:	0
Percent Similarity:	95.652	Percent Identity:	78.261

alignment_block:

US-09-684-016-48411 x H85462 ..

Align seg 1/1 to: H85462 from: 1 to: 258

5 TATTGTAATGTGAGATGCTTACACCTGATGACCTAATGCGCAATG 54
:::|||||
141 PheCYLSYSCYSGlUmetProTyrAsnProAspAspLeuMetValGlnCy 157
|||||
55 CGAGGGCTGCACCTGACTGG 73
|||||
157 sglUglUCySserGluTTP 163

seq_name: pir2:S69047

seq_documentation_block:

hypothetical protein YPL138c - Yeast (Saccharomyces cerevisiae)

C:Species: Saccharomyces cerevisiae

C:Date: 22-Aug-1996 #sequence_revision 06-Sep-1996 #text_change 05-Nov-1999

C:Accession: S69047

R:Hall, J.; DePaulo, T.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.
submitted to the EMBL Data Library, December 1995

A:Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.

A:Reference number: S69040

A:Accession: S69047

A:Molecule type: DNA

A:Residues: 1-353 <HAL>

A:Cross-references: EMBL:U43703; NID:91244769; PID:q1244777; GSPDB:GN00016; MIPS:YPL1

C:Genetics:

A:Gene: MIPS:YPL138c

A:Map position: 16L

alignment_scores:
 Quality: 66.50 Length: 23
 Ratio: 4.156 Gaps: 1
 Percent Similarity: 69.565 Percent Identity: 47.826

alignment_block:

US-09-684-016-48411 x S69047 ..

Align seg 1/1 to: S69047 from: 1 to: 353

```

5 TATTGTAATGTGAGATGCTTACACCTGATGACCTAATGGCGCAATG 54
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
24 TycysillecyslylsArgPro...AsptylcylglutleuMetValglcy 39
55 CGAGGCTGCACCTGACTGG 73
||||| ||||| ||||| ||||| ||||| ||||| |||||
39 SASpGlyCysAspAspTrp 45

```

seq_name: p1r2:T40419

seq_documentation_block:

m102 protein - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C:Accession: T40419
 R:McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Brown, S.; Harris, D.
 submitted to the EMBL Data Library, October 1999
 A:Reference number: Z21927
 A:Accession: T40419
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-329 <MCD>
 A:Cross-references: EMBL:AL121863; PIDN:CAB58404.1; GSPDB:GN00067; SPDB:SPBC4.05
 A:Experimental source: strain 972h; cosmid c4
 C:Genetics:
 A:Gene: SPDB:SPBC4.05
 A:Map position: 2

alignment_scores:
 Quality: 65.50 Length: 27
 Ratio: 4.367 Gaps: 1
 Percent Similarity: 55.556 Percent Identity: 44.444

alignment_block:

US-09-684-016-48411 x T40419 ..

Align seg 1/1 to: T40419 from: 1 to: 329

```

2 AGGATGTGAATGTGAGATGCTTACACCT.....GATGACT 42
||||| ||||| ||||| ||||| ||||| ||||| |||||
121 ArpPheCysIleCysAspTrpValTyrAsnProgluThrIuGlulGlyTh 137
43 AATGGCGCAATGGAGGCTGCACTGACTGG 73
||||| ||||| ||||| ||||| ||||| ||||| |||||
137 rMetPheGlnCysIleLeuGlyGluAspTrp 147

```

seq_name: p1r2:T41449

seq_documentation_block:

Probable phd finger transcription regulator - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 07-Dec-1999
 C:Accession: T41449
 R:Rieger, M.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.
 submitted to the EMBL Data Library, September 1998
 A:Reference number: Z21994
 A:Accession: T41449
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-424 <RIE>
 A:Cross-references: EMBL:AL031523; PIDN:CAA20664.1; GSPDB:GN00068; SPDB:SPCC594.05c

A:Experimental source: strain 972h; cosmid c594
 C:Genetics:
 A:Gene: SPDB:SPCC594.05c
 A:Map position: 3

alignment_scores:
 Quality: 62.50 Length: 25
 Ratio: 4.167 Gaps: 2
 Percent Similarity: 60.000 Percent Identity: 48.000

alignment_block:

US-09-684-016-48411 x T41449 ..

Align seg 1/1 to: T41449 from: 1 to: 424

```

5 TATTGTAATGTGAGATGCTTACACCTGATGAC.....CTAATGGC 48
||||| ||||| ||||| ||||| ||||| ||||| |||||
120 TycysillecysGln.....LysProAspAspIlySerTrpMetle 133
49 GCATGGGAGGCTGCACCTGACTGG 73
||||| ||||| ||||| ||||| ||||| ||||| |||||
133 uGlyCysAspGlyCysGluAspTrp 141

```

seq_name: p1r2:T21433

seq_documentation_block:

hypothetical protein F26H11.3a - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
 C:Accession: T21433
 R:Barlow, K.
 submitted to the EMBL Data Library, November 1996
 A:Reference number: Z19421
 A:Accession: T21433
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-405 <WIL>
 A:Cross-references: EMBL:Z81515; PIDN:CAB04198.1; GSPDB:GN00020; CESP:F26H11.3a
 A:Experimental source: clone F26H11
 C:Genetics:
 A:Gene: CESP:F26H11.3a
 A:Map position: 2
 A:introns: 11/1; 42/3; 183/2; 243/2; 287/1
 C:Superfamily: Bromodomain homology
 F:266-321/Domain: Bromodomain homology <BRO>

alignment_scores:
 Quality: 62.00 Length: 23
 Ratio: 4.429 Gaps: 0
 Percent Similarity: 60.870 Percent Identity: 34.783

alignment_block:

US-09-684-016-48411 x T21433 ..

Align seg 1/1 to: T21433 from: 1 to: 405

```

5 TATTGTAATGTGAGATGCTTACACCTGATGACCTAATGGCGCAATG 54
||||| ||||| ||||| ||||| ||||| ||||| |||||
172 TycysValcysGlnLysProTyrAspAspThrLysPheTyrValglcy 188
55 CGAGGCTGCACCTGACTGG 73
||||| ||||| ||||| ||||| ||||| ||||| |||||
188 SASpSerCysGlnGlyTrp 194

```

seq_name: p1r2:T21435

seq_documentation_block:

hypothetical protein F26H11.3c - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
 C:Accession: T21435

submitted to the EMBL Data Library, October 1998
A:Reference number: Z20293

A:Accession: T26971

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-368 <WTL>

A:Cross-references: EMBL:AL032657; PIDN:CAA21738.1; GSPDB:GN00019; CESP:y47H9C.2

A:Experimental source: clone y47H9C

C:Genetics:

A:Gene: CESP:y47H9C.2

A:Map position: 1

A:Introns: 83/2; 142/1; 194/1; 234/1; 262/2; 284/3; 339/2

alignment_scores:

Quality:	55.00	Length:	24
Ratio:	3.438	Gaps:	0
Percent Similarity:	66.667	Percent Identity:	45.833

alignment_block:

US-09-684-016-48411/rev x T26971 ..

Align seg 1/1 to: T26971 from: 1 to: 368

114 GAATTAATTTGAGATAGTATCCTTACATGTGAACCTTACAGTCAGT 65

||| ::::: ||||| |||||: ||:::

159 GUNHISValLysMethylsTyrCysThrThrCysArgLeuTyrArgProPr 175

64 GCAGCCCTCGCATTCGCCCAT 43

::: ||||| ||||| |||||

175 oArgCysSerHisCysAlaIle 182

seq_name: pir2:T32633

seq_documentation_block:

hypothetical protein F15E6.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T32633

R:Miller, N.; Stelleyes, L.; Bradshaw, H.; Keppler, D.

submitted to the EMBL Data Library, December 1997

A:Description: The sequence of C. elegans cosmid F15E6.

A:Reference number: Z21202

A:Accession: T32633

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1655 <WTL>

A:Cross-references: EMBL:AF038614; PIDN:AAB92063.1; GSPDB:GN00022; CESP:F15E6.1

A:Experimental source: strain Bristol N2; clone F15E6

C:Genetics:

A:Gene: CESP:F15E6.1

A:Map position: 4

A:Introns: 16/1; 68/1; 190/3; 301/2; 602/3; 1203/3; 1295/1; 1554/3; 1642/1

alignment_scores:

Quality:	54.50	Length:	22
Ratio:	4.192	Gaps:	1
Percent Similarity:	59.091	Percent Identity:	40.909

alignment_block:

US-09-684-016-48411 x T32633 ..

Align seg 1/1 to: T32633 from: 1 to: 1655

8 TGTAAATGTGAGATGCTTACACCCCTGATGACCTTAATGGCGCAATGGCA 57

||| ||| ||| ::::: ||| ::::: ||| ::::: |||

821 CysHIScysGlyMetAspHisGlyAspGlyAspThrIle...GlucySGL 836

58 GGGCTGCACCTGACTGG 73

||| ||| |||

836 uGlyCysLysThrTrp 841

seq_name: pir2:A31788

seq_documentation_block:

signal peptidase (EC 3.4.99.-) (SPC 22/22) - dog

C:Species: Canis lupus familiaris (dog)

C>Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 16-Feb-1997

C:Accession: A31788

R:Shelness, G.S.; Kanwar, Y.S.; Biobel, G.

J. Biol. Chem. 263, 17063-17070, 1988

A:Title: cDNA-derived primary structure of the glycoprotein component of canine micro-

A:Reference number: A31788; MUID:89034208

A:Accession: A31788

A:Molecule type: protein

A:Residues: 1-180 <SHE>

C:Keywords: glycoprotein; hydrolase; transmembrane protein

F;141/Binding site: carbohydrate (Asn) (covalent) #status predicted

alignment_scores:

Quality:	53.00	Length:	28
Ratio:	2.650	Gaps:	0
Percent Similarity:	71.429	Percent Identity:	42.857

alignment_block:

US-09-684-016-48411 x A31788 ..

Align seg 1/1 to: A31788 from: 1 to: 180

38 GACCTAATGCGCGCAATGCGAGCGCTGCACCTGCTGAATTCACATGT 87

||||: ||||: ||||: ||||| ||||| |||||

69 AsplIerhAlaAspLeuGlnAsnIlePheAspTrpAsnValLysGlnLe 85

88 AAGGTGATGATCTTCAATTTATCCACACAA 121

: ||||| ||||| ||||| ||||| |||||

85 uPheLeuTyrLeuSerAlaGluTyrSerThrLys 96

seq_name: pir2:P00542

seq_documentation_block:

polyprotein - sugarcane bacilliform virus (fragments)

C:Species: sugarcane bacilliform virus

C>Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 07-May-1999

C:Accession: P00542

R:Bouhida, M.; Lockhart, B.E.L.; Olszewski, N.E.

J. Gen. Virol. 74, 15-22, 1993

A:Title: An analysis of the complete sequence of a sugarcane bacilliform virus genome

A:Reference number: P00542; MUID:93139779

A:Accession: P00542

A:Molecule type: DNA

A:Residues: 1-46; 47-64; 65-101; 102-118; 119-132; 133-147; 148-161; 162-173; 174-185; 186-19

A:Cross-references: GB:M89923

C:Keywords: polyprotein

alignment_scores:

Quality:	53.00	Length:	19
Ratio:	3.786	Gaps:	0
Percent Similarity:	73.684	Percent Identity:	42.105

alignment_block:

US-09-684-016-48411 x P00542 ..

Align seg 1/1 to: P00542 from: 1 to: 199

8 TGTAAATGTGAGATGCTTACACCCCTGATGACCTTAATGGCGCAATGGCA 57

||||: ||| ::::: ||| ::::: ||| ::::: |||

47 CysArgcysTyrValcysGlySerProAspHisLeuMetLysAspCysLys 63

58 GGGCTGC 64

::::: |||

63 sSerCys 65

seq_name: pir2:B86073

A:Experimental source: strain 972h-; cosmid c645
C:Genetics:
A:Gene: SPDB:SPCC645.13
A:Map position: 3

alignment_scores:
Quality: 52.00 Length: 22
Ratio: 4.333 Gaps: 0
Percent Similarity: 54.545 Percent Identity: 36.364

alignment_block:
US-09-684-016-48411 x T41530 ..

Align seg 1/1 to: T41530 from: 1 to: 721

```

8 TGTAAATGTGATGCTTACACCGCTGATGCTTATGGCGCAATGCGA 57
   ||| |||::: ||| ||| ||| ||| ||| ||| ||| ||| |||
23 CysValCysLysSerGlnGluAspIleGlySerPThrTrpValGlnCysAs 39
58 GGGCGGACCTGACTGG 73
   ::||| |||
39 PGIYcysAspCysTrp 44

```

seq_name: p1r2:T06497

seq_documentation_block:

probable sucrose synthase (EC 2.4.1.13) 2 - garden pea

C:Species: Pisum sativum (garden pea)

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000

C:Accession: T06497

R:Buchner, P.

submitted to the EMBL Data Library, October 1997

A:Reference number: Z15720

A:Accession: T06497

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-809 <BUC>

A:Cross-references: EMBL:AJ001071; PIDN:CAA04512.1

A:Experimental source: cultivar Frisson

C:Superfamily: sucrose synthase; sucrose/sucrose-phosphate synthase homology

C:Keywords: glycosyltransferase; hexosyltransferase

F:282-757/Domain: sucrose/sucrose-phosphate synthase homology <SSPS>

alignment_scores:
Quality: 52.00 Length: 57
Ratio: 2.364 Gaps: 2
Percent Similarity: 38.596 Percent Identity: 26.316

alignment_block:
US-09-684-016-48411/rev x T06497 ..

Align seg 1/1 to: T06497 from: 1 to: 809

```

105 GAAGATAGATATCCTTACCTGACG..... 79
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
457 GluAspLysTrpHisPheSerCysGlnPheThrAlaAspLeuIleAlaMe 473
78 ..... ACTTACCACTGAGTG..... 64
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
473 LAsnAsnAlaAspPheIleIleThrSerThrTyrGlnGluIleAlaGlyT 490
63 ..... CAGCCCTGCGATTGCGCCATTAGTGCATCA 34
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
490 hTrLysAsnThrIleGlyGlnTyrGlnSerHisThrAlaPheThrLeuPro 506
33 GGGTTGTAGGACATCCACAT 13
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
507 GlyLeuTyrArgValValHis 513

```

seq_name: p1r2:T16487

seq_documentation_block:
hypothetical protein F56F10.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 04-Mar-2000
C:Accession: T16487
R:Bradshaw, H.

submitted to the EMBL Data Library, March 1996
A:Description: The sequence of C. elegans cosmid F56F10.
A:Reference number: Z18521

A:Accession: T16487

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-269 <BRAS>

A:Cross-references: EMBL:U51993; NID:g1688050; PID:g1255790; PIDN:AA836851.1; GSPDB:(

A:Experimental source: strain Bristol NZ; clone F56F10

C:Genetics:

A:Gene: CESP:F56F10.2

A:Map position: X

A:Introns: 21/3; 45/2; 100/3; 147/3; 177/3; 232/3

C:Superfamily: Caenorhabditis elegans hypothetical protein F56F10.2

alignment_scores:
Quality: 51.50 Length: 37
Ratio: 2.060 Gaps: 1
Percent Similarity: 67.568 Percent Identity: 32.432

alignment_block:
US-09-684-016-48411/rev x T16487 ..

Align seg 1/1 to: T16487 from: 1 to: 269

```

119 GTGTGGATTAATTCAGAGATA...AGTATCACCTTACATGTAACACTTAC 73
   ::||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
221 IleTrpSerLysPheSerAlaAspSerIlePheLeuPheGlnGluLeuVal 237
72 CAGTCAGTCAGACCCCTGCTGCGCCATTGCGCCATTAGTCATCAGGTTGTAGG 23
   ::||| ||| ::||| ::||| ::||| ::||| ::||| ::|||
237 lLysGlnAsnSerAspThrValThrAspPheGlyLeuGlnLucySerP 254
22 CATCTCACATT 12
   ||||| |||||
254 roSerHisLeu 257

```

seq_name: p1r2:A40084

seq_documentation_block:

exogastrola-inducing protein (EGIP) precursor - sea urchin (Strongylocentrotus purpur

N:Alternate names: epidermal growth factor-related protein

N:Contains: EGIP peptide A; EGIP peptide B; EGIP peptide C; EGIP peptide D

C:Species: Strongylocentrotus purpuratus (purple urchin)

C:Date: 20-Mar-1992 #sequence_revision 15-Aug-1997 #text_change 11-Jan-2000

C:Accession: A40084

R:Yang, Q.; Angerer, L.M.; Angerer, R.C.

Science 246, 806-808, 1989

A:Title: Unusual pattern of accumulation of mRNA encoding EGF-related protein in sea

A:Reference number: A40084; M0ID:90049203

A:Accession: A40084

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-325 <YAN>

A:Cross-references: GB:M29004; NID:g161463; PID:g161464

A:Note: the publication does not report the complete sequence

C:Superfamily: unassigned EGF-related proteins; EGF homology

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-325/Product: exogastrola-inducing protein #status predicted <EGIP>

F:47-104/Product: EGIP peptide C #status predicted <EGIC>

F:52-90/Domain: EGF homology <EGI>

F:106-156/Product: EGIP peptide D #status predicted <EGID>

F:111-151/Domain: EGF homology <EG2>

F:177-228/Product: EGIP peptide A #status predicted <EGIA>

F:182-223/Domain: EGF homology <EG3>

F:250-309/Product: EGIP peptide B #status predicted <EGIB>

F:256-292/Domain: EGF homology <EG4>
F:52-65,59-75,77-90,111-124,118-137,139-151,182-195,189-209,211-223,256-269,263-278,280-

alignment_scores:

Quality: 51.50 Length: 14
Ratio: 4.682 Gaps: 1
Percent Similarity: 78.571 Percent Identity: 64.286

alignment_block:

US-09-684-016-48411 x A40084 ..

Align seg 1/1 to: A40084 from: 1 to: 325

5 TATGTGAATGTGAGATGCTTACAC.....CCTGAT 37
|||||
74 TYRCYSGLYCYSGLUETPROPHEARGVALGYLLEPROASP 87

seq_name: p1r2:T38168

seq_documentation_block:

hypothetical protein SPAC22E12.11c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
R:Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rastandream, M.A.; Walsh, S.V.
Submitted to the EMBL Data Library, March 1996
A:Accession: T38168
A:Reference number: 221775
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-859 <DEV>
A:Cross-references: EMBL:Z70043; NID:g1220275; PIDN:CAA3898.1; GSPDB:GN00066; SPDB:SPAC
A:Experimental source: strain 972h; cosmid c22E12
C:Genetics:
A:Gene: SPDB:SPAC22E12.11c
A:Map position: 1

alignment_scores:

Quality: 51.50 Length: 22
Ratio: 3.679 Gaps: 1
Percent Similarity: 63.636 Percent Identity: 40.909

alignment_block:

US-09-684-016-48411 x T38168 ..

Align seg 1/1 to: T38168 from: 1 to: 859

11 AATGTGAATGCTTACACCTGATGAC..CTAATGGCGAATGCGA 57
:::|||||
5 ATGCVSVALCYSTROPHEGLUASPASPGLYPHERTRILEGNCYSGL 21
58 GGGCTGCACTGACTGG 73
|:::|||||
21 usercysgluvaltrp 26

seq_name: p1r2:D96660

seq_documentation_block:

protein F2K11.14 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.E.; Hughes, B.; Hultzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, Z.A.; Lueros, J.S.; Malt, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: D96660

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-1518 <STO>

A:Cross-references: GB:AE005173; NID:96633837; PIDN:AA19696.1; GSPDB:GN00141

C:Genetics:

A:Gene: F2K11.14

A:Map position: 1

alignment_scores:

Quality: 51.50 Length: 23
Ratio: 3.433 Gaps: 1
Percent Similarity: 65.217 Percent Identity: 39.130

alignment_block:

US-09-684-016-48411 x D96660 ..

Align seg 1/1 to: D96660 from: 1 to: 1518

5 TATGTGAATGTGAGATGCTTACACCTGATGACCTAATGCGCAATG 54
:::|||||
1411 HISCYSLIECYSLIEULYSPROTIFYRANSEARYSER..Metvalsercy 1426
55 CGAGGCTGCGACTGCTGCTG 73
|:::|||||
1426 SSERGLNCSGLYGLUTRP 1432

seq_name: p1r2:T26611

seq_documentation_block:

hypothetical protein Y37A1B.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T26611
R:McMurray, A.
Submitted to the EMBL Data Library, June 1998
A:Reference number: Z20245
A:Accession: T26611
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-95 <ML>
A:Cross-references: EMBL:AL023835; PIDN:CAA19489.1; GSPDB:GN00022; CESP:Y37A1B.4
A:Experimental source: clone Y37A1B
C:Genetics:
A:Gene: CESP:Y37A1B.4
A:Map position: 4
A:introns: 46/2; 84/3

alignment_scores:

Quality: 51.00 Length: 36
Ratio: 2.833 Gaps: 1
Percent Similarity: 50.000 Percent Identity: 30.556

alignment_block:

US-09-684-016-48411/rev x T26611 ..

Align seg 1/1 to: T26611 from: 1 to: 95

122 GTTGTGTAATGAATGAAGATAGTACCTTACATGATGAACTTAC 73
:::|||||
46 ILELETRPASNINLEUVALMETILE..... 55
72 CACTGAGTCAGCCCTGCGCACTTGGCCATTAGGTCATCAGGGTTGACG 23
|||||
56GLNCYASPGLYASNLEULEUPROGLNYSHISGLNGLYPHETRYA 71
22 CATCTCAC 15
|||
71 rphphls 73

```

seq_name: p1r2:F64828

seq_documentation_block:
  anaerobic dimethylsulfoxide reductase (EC 1.-.-.) chain B - Escherichia coli
  C:Species: Escherichia coli
  C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 11-Jun-1999
  C:Accession: F64828; S03786; 090169
  R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
    .A.; Rose, D.; Mau, B.; Shao, Y.
  Science 277, 1453-1462, 1997
  A:Title: The complete genome sequence of Escherichia coli K-12.
  A:Reference number: A64720; MUID:97426617
  A:Accession: F64828
  A:Status: nucleic acid sequence not shown; translation not shown
  A:Molecule type: DNA
  A:Residues: 1-205 <GBAT>
  A:Cross-references: GB:AE000191; GB:U00096; NID:g1787115; PIDN:AAC73981.1; PID:g1787122.
  A:Experimental source: strain K-12, substrain MG1655
  R:Billous, P.T.; Cole, S.T.; Anderson, W.F.; Welner, J.H.
  Mol. Microbiol. 2, 785-795, 1988
  A:Title: Nucleotide sequence of the dmsABC operon encoding the anaerobic dimethylsulphox
    A:Reference number: S03784; MUID:89096500
  A:Molecule type: DNA
  A:Accession: S03786
  A:Cross-references: GB:J03412; NID:g145754; PIDN:AA83844.1; PID:g145756
  A:Note: part of this sequence, including the amino end of the mature protein, was confli
    C:Genetics:
    A:Gene: dmsB
    A:Map position: 20 min
    A:Complex: heterotrimer; chains A, B, and C
    C:Function:
    A:Description: terminal reductase during anaerobic growth on various sulfoxide and N-oxi
    A:Note: chain A binds molylidoprotein, chain B is an electron transfer protein, and chain
    C:Keywords: 4Fe-4S; heterotrimer; iron-sulfur protein; metalloprotein; oxidoreductase
    F:2-205/Product: dimethylsulfoxide reductase chain B #status experimental <MAT>
    F:7-87/Domain: ferredoxin 2(4Fe-4S) homology <FER1>
    F:92-153/Domain: ferredoxin 2(4Fe-4S) homology <FER2>
    F:11,12,20,79/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted
    F:24,67,70,75/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted
    F:99,102,105,145/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted
    F:109,126,129,141/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

alignment_scores:
  Quality: 51.00 Length: 29
  Ratio: 3.400 Gaps: 1
  Percent Similarity: 51.724 Percent Identity: 34.483

alignment_block:
  US-09-684-016-48411 x F64828 ..
  Align seg 1/1 to: F64828 from: 1 to: 205

  2 AGGTATTTGTAATGTGAGATGCTTACAAACCTGATGACCTA..... 43
  ||||||| |||||||
  103 ArgTYrCYSHisMetalAcysProTYrGLyAlaProGlnTYrAsnGLuTh 119
  44 .....ATGGCGCAATGCGAGGCGCTGCACTGAC 70
  |||:|||||:||||| ||
  119 rlysglyHismethrlyscysAspGlyCYsTYrAsp 131

seq_name: p1r2:B85615

seq_documentation_block:
  hypothetical protein dmsB [imported] - Escherichia coli (strain O157:H7)
  C:Species: Escherichia coli
  C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
  C:Accession: B85615
  R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
    .A.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,

```

```

Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:1120551
A:Accession: B85615
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-205 <STO>
A:Cross-references: GB:AB005174; NTD:g12514060; PIDN:AA655382.1; GSPDB:GN00145; UMGF
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: dmsB

alignment_scores:
  Quality: 51.00      Length: 29
  Ratio: 3.400      Gaps: 1
  Percent Similarity: 51.724      Percent Identity: 34.483

alignment_block:
US-09-684-016-48411 x T21654 ..

Align seg 1/1 to: B85615 from: 1 to: 205

2 AGGTATGTAATGTGAGATGCCTTACAACCGATGACCTA..... 43
|||||  |||||||  |||||||  ::
103 ATGTTCySHisMetAlaCysProTyrGlyAlaProGlnTyrAsnGluThr 119
44 .....ATGGCGCAATGCGAGGCGCTGCACCTGAC 70
|||||  |||||||  |||
119 rlysglyHisMetThrLysCysAspGlyCysTyrAsp 131

seq_name: p1r2:T21654

seq_documentation_block:
hypothetical protein F32D8.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T21654
R:Wilkinson, J.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19454
A:Accession: T21654
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-245 <WILL>
A:Cross-references: EMBL:Z74031; PIDN:CAA98455.1; GSPDB:GN00023; CESP:F32D8.3
A:Experimental source: Clone F32D8
C:Genetics:
A:Gene: CESP:F32D8.3
A:Map position: 5
A:Introns: 61/1; 83/3
C:Superfamily: Caenorhabditis elegans hypothetical protein F32D8.3

alignment_scores:
  Quality: 51.00      Length: 23
  Ratio: 3.643      Gaps: 2
  Percent Similarity: 60.870      Percent Identity: 47.826

alignment_block:
US-09-684-016-48411 x T21654 ..

Align seg 1/1 to: T21654 from: 1 to: 245

5 TTTTGAATGTGAGATGCCT...TACAACCGATGACCTAATGGGCA 51
:::|||||  |||||||  |||||||  :::::
71 HisCysAlaCysGlnSerThrCysAsnAspProAspProTyrCysSerIy 87
52 ATGCGAG...GGCTGCACT 67
:|||||  |||||||  |||
87 scysglnProGlyCysThr 93

seq_name: p1r2:T04224

```

seq_documentation_block:
 hypothetical protein T5C23.200 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 24-Nov-1999
 C:Accession: T04224
 R:Bayan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft, submitted to the Protein Sequence Database, March 1999
 A:Reference number: 215261
 A:Accession: T04224
 A:Molecule type: DNA
 A:Residues: 1-386 <BEV>
 A:Cross-references: EMBL:AL049500
 A:Experimental source: cultivar Columbia; BAC clone T5C23
 C:Genetics:
 A:Map position: 4
 A:Note: T5C23.200
 C:Superfamily: caffeoyl-CoA 3-O-methyltransferase

alignment_scores:
 Quality: 51.00 Length: 16
 Ratio: 4.250 Gaps: 0
 Percent Similarity: 75.000 Percent Identity: 56.250

alignment_block:
 US-09-684-016-48411 x T04224 ..

Align seg 1/1 to: T04224 from: 1 to: 396

```

14 TGTGAGATGCGCTTACACCTGATGACCTATGCGGCAATGCGAGGC 61
   |||:|||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
10 CysAsmMetProTyRleuProAspAspLeuLeuAsnIleLeuGly 25

```

seq_name: pir2:A35291

seq_documentation_block:
 adenylosuccinate lyase (EC 4.3.2.2) - chicken
 C:Species: Gallus gallus (chicken)
 C:Date: 17-Aug-1990 #sequence_revision 17-Aug-1990 #text_change 21-Jul-2000
 C:Accession: A35291
 R:Alml, J.; Badylak, J.; Williams, J.; Chen, Z.; Zalkin, H.; Dixon, J.E.
 J. Biol. Chem. 265, 9011-9014, 1990
 A:Title: Cloning of a cDNA encoding adenylosuccinate lyase by functional complementation
 A:Reference number: A35291; MUID:90264380
 A:Accession: A35291
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-459 <AIM>
 A:Cross-references: GB:M37901; NID:9211094; PIDN:AAA48574.1; PID:9211095
 C:Superfamily: fumarate hydratase
 C:Keywords: amidine-lyase; carbon-nitrogen lyase

alignment_scores:
 Quality: 51.00 Length: 33
 Ratio: 2.550 Gaps: 0
 Percent Similarity: 60.606 Percent Identity: 27.273

alignment_block:
 US-09-684-016-48411/rev x A35291 ..

Align seg 1/1 to: A35291 from: 1 to: 459

```

110 AAATGAAAGTAAAGTATCAGCTTACATGTAACCTTACACGATGTCAG 61
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
58 LysLeuArgHisAspValMetAlaHisValHisThrPheAlaHisCys 74
60 CCCTGCGATTCGCGCATTAAGGTCATGAGGTCGATTAAGGTCATCATT 12
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
74 sProLysAlaAlaAlaIleIleHisLeuGlyAlaThrSerCysTyrVal 90

```

seq_name: pir2:F71844

seq_documentation_block:
 ABC transporter, ATP-binding protein - Helicobacter pylori (strain J99)
 C:Species: Helicobacter pylori
 A:Variety: strain J99
 C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 17-Mar-2000
 C:Accession: F71844
 R:Alm, R.A.; Ling, L.S.L.; Molr, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.
 Nature 397, 176-180, 1999
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p
 A:Reference number: A71800; MUID:99120557
 A:Accession: F71844
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-578 <ARN>
 A:Cross-references: GB:AE001541; GB:AE001439; NID:94155724; PIDN:AD06712.1; PID:9415
 A:Experimental source: strain J99
 C:Genetics:
 A:Gene: jhp1129
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homolog
 F:381-573/Domain: ATP-binding cassette homolog <ABC>

alignment_scores:
 Quality: 51.00 Length: 31
 Ratio: 2.684 Gaps: 0
 Percent Similarity: 61.290 Percent Identity: 38.710

alignment_block:
 US-09-684-016-48411/rev x F71844 ..

Align seg 1/1 to: F71844 from: 1 to: 578

```

110 AAATGAAAGTAAAGTATCAGCTTACATGTAACCTTACACGATGTCAG 61
   ||| ||| :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
375 LysSerLysHisProValLeuLysAsnPhaHisLeuThrIleIleLys 391
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
60 CCCTGCGATTCGCGCATTAAGGTCATGAGGTCGATTAAGGTCATCT 18
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
391 yGlnLysIleAlaLeuIleGlyHisSerGlyCysGlyLysSer 405

```

seq_name: pir2:H86457

seq_documentation_block:
 hypothetical protein AAG21600.1 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: H86457
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
 ansen, N.F.; Hughes, B.; Hultzer, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matli, R.; Marzia
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, C.M.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: H86457
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-697 <STO>
 A:Cross-references: GB:AE005172; NID:910645488; PIDN:AAG21600.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

alignment_scores:
 Quality: 51.00 Length: 28
 Ratio: 3.643 Gaps: 1
 Percent Similarity: 50.000 Percent Identity: 35.714

alignment_block:

US-09-684-016-48411 x H86457 ..

Align seg 1/1 to: H86457 from: 1 to: 697

```

      8 TGTAATGTGAGATGCCCTTACACCCCTGATGACCTTATGCGCATTCGCA 57
      |||||||
      606 CyslyscysglYThrlyAspAspAspAspGlyGluArgMetLeuAlaCysAs 622
      |||||||
      58 GGGCTGACACTGATGCTGTAAGTTTCACATGTAAG 91
      |||||||
      622 pGlyCysGlyValTrp.....HisHisThrArg 631

```

seq_name: p1r2:T15670

seq_documentation_block:

hypothetical protein C27H5.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

C:Accession: T15670

R:Pauley, A.

submitted to the EMBL Data Library, July 1995

A:Description: The sequence of C. elegans cosmid C27H5.

A:Reference number: Z18386

A:Accession: T15670

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1332 <PAU>

A:Cross-references: EMBL:U14635; NID:9540265; PID:9540271; PIDN:MAC46659.1; CESP:C27H5.5

A:Experimental source: strain Bristol N2

C:Genetics:

A:Gene: CESP:C27H5.5

A:Introns: 27/2; 93/3; 203/1; 336/1; 358/2; 394/3; 473/2; 489/3; 531/3; 574/2; 636/3; 67

alignment_scores:

Quality:	50.50	Length:	36
Ratio:	2.196	Gaps:	3
Percent Similarity:	63.889	Percent Identity:	36.111

alignment_block:

US-09-684-016-48411/rev x T15670 ..

Align seg 1/1 to: T15670 from: 1 to: 1332

```

      95 ATCACTTACATGTGAAGTAACCAAGTCAGTCAGACCCCTCGCATTCGCG.. 48
      ::|||
      176 LeuThrLeuHisLeu.....AlaCysSerMetHisLeuSerAr 188
      |||||
      47 ....CCATTAGTCATCAGGTTGTAG.....GCATCTGCATTT 11
      |||||
      188 gTyrPProMetAspHisGlnAsnCysGluIleAlaPheAlaSerYrAlaT 205
      |||||
      10 ACAATACC 3
      |||||
      205 YrThrThr 207

```

seq_name: p1r2:S22412

seq_documentation_block:

signal peptidase (EC 3.4.99.-), microsomal - chicken

C:Species: Gallus gallus (chicken)

C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 21-Jul-2000

C:Accession: S22412; S40018; S15551

R:Newsome, A.L.; McLean, J.W.; Lively, M.O.

Biochem. J. 282, 447-452, 1992

A:Title: Molecular cloning of a cDNA encoding the glycoprotein of hen oviduct microsomal

A:Reference number: S22412; MUID:92189580

A:Accession: S22412

A:Molecule type: mRNA

A:Residues: 1-180 <MEM1>

A:Cross-references: EMBL:X60795; NID:963421; PIDN:CAA43208.1; PID:963422

A:Accession: S40018

A:Molecule type: protein

A:Residues: 1-15;36-45;50-58;76-83;97-118;141-160 <MEM2>

C:Keywords: glycoprotein; hydrolase; transmembrane protein

F:1-180/Product: signal peptidase #status experimental <MAT>

F:141/Binding site: carbohydrate (Asn) (covalent) #status experimental

alignment_scores:

Quality:	50.00	Length:	28
Ratio:	2.500	Gaps:	0
Percent Similarity:	71.429	Percent Identity:	39.286

alignment_block:

US-09-684-016-48411 x S22412 ..

Align seg 1/1 to: S22412 from: 1 to: 180

```

      38 GACCTAATGGCGCAATGCGAGGCTGCACCTGACGTGTAAGTTTCACATGT 87
      |||||
      69 AspiIeThrAlaAspLeuGlnSerIlePheAspTrpAsnValGlnIle 85
      |||||
      88 AAGGTGATCACTTATCTTCATTTATTCACACAA 121
      |||||
      85 uPheLeuTyrLeuSerAlaGluTyrSerThrlys 96

```

seq_name: p1r2:S61299

seq_documentation_block:

lipopolysaccharide heptosyltransferase (EC 2.4.99.-) II rfaI [similarity] - Neisser

C:Species: Neisseria gonorrhoeae

C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 21-Jul-2000

C:Accession: S61299

R:Schwan, E.T.; Robertson, B.D.; Brade, H.; van Putten, J.P.M.

Mol. Microbiol. 15, 267-275, 1995

A:Title: Gonococcal rfaI mutants express Rd(2) chemotype LPS and do not enter epithel

A:Reference number: S61299; MUID:95264913

A:Accession: S61299

A:Status: preliminary; nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-336 <SCH>

A:Cross-references: EMBL:Z27141; NID:9599919; PIDN:CAA85504.1; PID:9599920

A:Note: the sequence of residues 1-2 are not shown in this paper

C:Superfamily: ADP-heptose-LPS heptosyltransferase II

C:Keywords: glycosyltransferase

alignment_scores:

Quality:	50.00	Length:	29
Ratio:	2.632	Gaps:	2
Percent Similarity:	65.517	Percent Identity:	41.379

alignment_block:

US-09-684-016-48411/rev x S61299 ..

Align seg 1/1 to: S61299 from: 1 to: 336

```

      113 AATAAATTGAAGATGATATCACCTTACATGTGAAGACTTACACGACGTG 64
      ::|||
      293 AsparGalaLysIle...ValSerLeuHisLeuGlu...CysSerProCys 307
      |||
      63 CAGCCCTGCGCATTCGCGCATTCAGGAGGTTGT 27
      |||
      307 sPheLysArgGluCysProLeuGlyHisThrAspCys 319

```

seq_name: p1r2:B81869

seq_documentation_block:

lipopolysaccharide heptosyltransferase (EC 2.4.99.-) II NKA1727 [similarity] - Neiss

C:Species: Neisseria meningitidis

C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C:Accession: B81869

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; M

: Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.
A:Reference number: AB1775; MUID:20222556
A:Accession: B81869
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-336 <PAR>
A:Cross-references: GB:AL162756; GB:AL157959; NID:g7380091; PIDN:CAB84955.1; PID:g738036
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: rfaF; NMAL727
C:Superfamily: ADP-heptose--LPS heptosyltransferase II
C:Keywords: glycosyltransferase

alignment_scores:
Quality: 50.00 Length: 29
Ratio: 2.632 Gaps: 2
Percent Similarity: 65.517 Percent Identity: 41.379

alignment_block:
US-09-684-016-48411/rev x B81869 ..

Align seg 1/1 to: B81869 from: 1 to: 336

```
113 AATTAATTGAAGATAGTATCACCTTACATGTCGAAGCTTACACAGTCAGTG 64
      ::::: ||||| :::::|||||::: || ||
293 AspaRgAlaIaIaIle...ValSerLeuHISLeuGlu...CysSerProCy 307
      ||| :::||||| |||
63 CAGCCCTCGCATTCGCGCATTAGTCATCAGGCTGT 27
      ||| :::||||| |||
307 spheLysaRgGluGlyCysProLeuGlyHISLtnrAspCys 319
```

OM of: US-09-684-016-48411 to: SwissProt_39.* out_format : pfs

Date: Mar 19, 2002 9:34 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODE=frame+npz.model -DBV=tblh
-Q/cgn2.1/USPTO.spool/US09684016/runat.19032002.084440.20960/app.query.fasta.1.182
-DB=SwissProt_39 -QPM=fastan -SUFFIX=std.rsp -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCGAPOP=6.000 -FCGAPEXT=7.000 -YGAPEXT=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -STAFF=1 -MATRIX=blomsurf62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR.SCORE=pct
-THR.MAX=100 -THR.MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFM=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09684016.@CGN1.1.28 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-684-016-48411
Query length: 123
Database: SwissProt_39.*
Database sequences: 100059
Database length: 36664827
Search time (sec): 26.420000
```

score_list:

Sequence	Stnd Orig	ZScore	EScore Len	Document
SwissProt_39:ML02_SCHPO	65.50	161.74	0.0747	329 009329 schizosaccharomyces
SwissProt_39:MOBA_NEIMA	59.00	148.38	0.7099	192 09145 neisseria meningitidis
SwissProt_39:FTS2_BABBA	56.50	132.34	1.80	592 03114 bartonella bacilliformis
SwissProt_39:CKR_XENIA	56.50	131.97	1.81	619 09175 xenopus laevis (afri
SwissProt_39:OPLA_RAT	56.00	124.61	2.23	1288 P97608 rattus norvegicus (rat
SwissProt_39:H1P1_HUMAN	54.00	121.26	4.44	995 000291 homo sapiens (human)
SwissProt_39:MOBA_ECO57	53.00	132.52	5.79	180 P12280 canis familiaris (dog
SwissProt_39:MOBA_ECO57	52.00	129.17	8.26	194 P58221 escherichia coli o157
SwissProt_39:MOBA_ECO57	52.00	129.17	8.26	194 P32173 escherichia coli o157
SwissProt_39:HEX2_STEPU	51.50	123.59	10.09	325 P15216 strongyloides
SwissProt_39:HEX2_STEPU	51.50	115.63	10.59	859 P10362 schizosaccharomyces
SwissProt_39:HEX2_STEPU	51.00	131.62	11.36	103 06594 canine adenovirus 1
SwissProt_39:HEX2_STEPU	51.00	126.03	11.75	204 P18776 escherichia coli o157
SwissProt_39:HEX2_STEPU	51.00	119.39	12.23	459 P2165 gallus gallus (chick)
SwissProt_39:HEX2_STEPU	50.50	128.23	13.70	132 P26498 human adenovirus type
SwissProt_39:HEX2_STEPU	50.00	124.32	16.58	180 P26887 gallus gallus (chick)
SwissProt_39:HEX2_STEPU	50.00	117.25	17.30	427 P08138 homo sapiens (human)
SwissProt_39:HEX2_STEPU	50.00	115.05	17.53	559 092128 mus musculus (mouse)
SwissProt_39:HEX2_STEPU	50.00	114.93	17.54	567 094189 homo sapiens (human)
SwissProt_39:HEX2_STEPU	50.00	114.46	17.60	601 095572 strongyloides
SwissProt_39:HEX2_STEPU	50.00	112.47	17.81	766 042652 beta vulgaris (sugar
SwissProt_39:HEX2_STEPU	50.00	105.95	18.52	1700 003376 chironomus tentans
SwissProt_39:HEX2_STEPU	50.00	105.84	18.54	1722 P29375 homo sapiens (human)
SwissProt_39:HEX2_STEPU	49.50	120.14	20.09	254 P21640 pseudomonas denitrifi
SwissProt_39:HEX2_STEPU	49.50	105.06	22.01	1603 007092 homo sapiens (human)
SwissProt_39:HEX2_STEPU	49.00	112.30	24.89	560 P42948 saccharomyces cerevis
SwissProt_39:HEX2_STEPU	49.00	109.02	25.39	836 P13724 homo sapiens (human)
SwissProt_39:HEX2_STEPU	48.50	112.28	29.42	475 092128 mus musculus (mouse)
SwissProt_39:HEX2_STEPU	48.50	108.96	30.02	713 P20299 mus musculus (mouse)
SwissProt_39:HEX2_STEPU	48.50	108.34	30.13	769 P32292 bos taurus (bovine)
SwissProt_39:HEX2_STEPU	48.50	107.70	30.25	832 P25611 saccharomyces cerevis
SwissProt_39:HEX2_STEPU	48.00	120.03	33.17	156 093564 caenorhabditis elegans
SwissProt_39:HEX2_STEPU	48.00	117.79	33.62	205 P45003 haemophilus influenza
SwissProt_39:HEX2_STEPU	48.00	115.85	34.02	260 091085 melospiza gallopavo
SwissProt_39:HEX2_STEPU	48.00	114.90	34.22	292 016655 homo sapiens (human)
SwissProt_39:HEX2_STEPU	48.00	113.46	34.52	348 025331 manduca sexta (tobac
SwissProt_39:HEX2_STEPU	48.00	112.57	34.70	388 094963 homo sapiens (human)
SwissProt_39:HEX2_STEPU	48.00	112.57	34.70	388 094963 homo sapiens (human)
SwissProt_39:HEX2_STEPU	48.00	112.00	34.82	416 P18519 gallus gallus (chick)

```
SwissProt_39:NGFR_RAT + 48.00 111.83 34.86 425 | P07174 rattus norvegicus (
SwissProt_39:GAC3_HUMAN - 48.00 111.06 35.02 467 | 099298 homo sapiens (hum
SwissProt_39:MM17_HUMAN - 48.00 108.92 35.48 606 | 099129 homo sapiens (hum
SwissProt_39:APC2_CAVPO + 47.50 122.30 38.66 100 | P27916 cavia porcellus (
SwissProt_39:EGIP_ANTCR + 47.50 112.66 40.99 325 | P15217 anthracidis cras
```

seq_name: SwissProt_39:ML02_SCHPO

seq_documentation_block:

```
ID ML02_SCHPO STANDARD: PRT: 329 AA.
AC 009329;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN ML02
GN ML02 OR SPB4.05.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=97128260; PubMed=8972853;
RA Javerz J.-P., Cranston G., Allshire R.A.;
RT "Fission yeast genes which disrupt mitotic chromosome segregation
RT when overexpressed.";
RL Nucleic Acids Res. 24:4676-4683(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA McDonnell R.C., Rajandream M.A., Barrell B.G., Brown S., Harris D.;
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RL -! FUNCTION: NOT KNOWN, INTERFERE WITH MITOTIC CHROMOSOME SEGREGATION
WHEN OVEREXPRESSED.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
CC EMBL: L42550; AAB41271.1; -.
CC EMBL: AL121863; CAB58404.1; -.
CC InterPro: IPR001965; PHD.
CC InterPro: IPR001126; ZNF_UBR1.
CC Pfam: PF00628; PHD. 1.
CC Pfam: PF02207; ZF-UBR1. 1.
CC SMART: SM00249; PHD. 1.
CC SMART: SM00396; ZNF_UBR1. 1.
CC SQUAD: 329 AA; 38078 MW; 42544382917195F CRC64;
SQ
```

alignment_scores:

```
Quality: 65.50 Length: 27
Ratio: 4.367 Gaps: 1
Percent similarity: 55.556 Percent identity: 44.444
```

alignment_block:

US-09-684-016-48411 x ML02_SCHPO ..

Align seg 1/1 to: ML02_SCHPO from: 1 to: 329

```
2 AGGATGTAATGAGAGAGCTTACACCT.....GATGACCT 42
|||||:|||||:|||||:|||||:|||||:
121 ArgpneCysIleCysAspThrValIlyrnsProGluTrgIuGluGlyTh 137
43 AATGGCGAATGCGAGGCGTCACACTGG 73
||| ||||| ||| |||||
137 MetheGlnCysIleLeuGlyGluAspTrp 147
```

```

seq_name: SwissProt_39:MOBA_NEIMA
seq_documentation_block:
ID MOBA_NEIMA STANDARD: PRT; 192 AA.
AC O9JUA5:
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE PROBABLE MOLYBDOPTERIN-GUANINE DINUCLEOTIDE BIOSYNTHESIS PROTEIN A.
GN MOBA OR NMA1417.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=63699;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-22491 / SEROTYPE 4A;
MEDLINE=2022556; PubMed=10761919;
RA Parthill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
meningitidis 22491."
RL Nature 404:502-506(2000).
CC -1- FUNCTION: LINKS A GUANOSINE 5'-PHOSPHATE TO MOLYBDOPTERIN (MPT)
FORMING MOLYBDOPTERIN GUANINE DINUCLEOTIDE (MGD) (BY SIMILARITY).
CC -1- PATHWAY: MOLYBDENUM COFACTOR BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE MOBA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AL162755; CAB84657.1;
KW Molybdenum cofactor biosynthesis; GTP-binding; Complete proteome.
SQ SEQUENCE 192 AA; 21803 MW; 005C495C6C64D290 CRC64;

alignment_scores:
Quality: 59.00 Length: 15
Ratio: 4.538 Gaps: 0
Percent Similarity: 86.667 Percent Identity: 66.667

alignment_block:
US-09-684-016-48411 x MOBA_NEIMA ..

Align seg 1/1 to: MOBA_NEIMA from: 1 to: 192

14 TGTGAGATGCTTACACCTGATGACCTAATGGCGCAATGCGAG 58
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
100 CysAspMetProTyrLeuProAspAspLeuValAlaArgPheGlu 114

seq_name: SwissProt_39:FTSZ_BARBA
seq_documentation_block:
ID FTSZ_BARBA STANDARD: PRT; 592 AA.
AC O31314:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE CELL DIVISION PROTEIN FTSZ (75 KDA ANTIGEN).
GN FTSZ.
OS Bartonella bacilliformis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bartonellaceae; Bartonella.

```

```

OX NCBI_TaxID=774;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-KC584;
MEDLINE=97369823; PubMed=9226264;
RA Padmalayam I., Anderson B., Kiron M., Kelly T., Baumstark B.;
RT "The 75-kilodalton antigen of Bartonella bacilliformis is a
structural homolog of the cell division protein ftsZ."
RL J. Bacteriol. 179:4545-4552(1997).
RN [2]
RP REVISIONS.
RA Kelly T.;
RL Submitted (APR-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: THIS PROTEIN IS ESSENTIAL TO THE CELL-DIVISION PROCESS.
ITS SEEMS TO ASSEMBLE INTO A DYNAMIC RING ON THE INNER SURFACE OF
THE CYTOPLASMIC MEMBRANE AT THE PLACE WHERE DIVISION WILL OCCUR,
AND THE FORMATION OF THE RING IS THE SIGNAL FOR SEPTATION TO
BEGIN. BINDS TO AND HYDROLYZES GTP (BY SIMILARITY).
CC -1- SUBUNIT: AGGREGATE TO FORM A RING-LIKE STRUCTURE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC; ASSEMBLE AT THE INNER SURFACE
OF THE CYTOPLASMIC MEMBRANE (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF007266; AAC15082.1;
DR HSSP: Q57816; FTSZ.
DR InterPro: IPR000158; FtsZ.
DR InterPro: IPR003008; Tubulin_FtsZ.
DR Pfam: PF00091; tubulin_1.
DR PRINTS: PR00423; CELDIVISFTSZ.
DR PROSITE: PS01134; FTSZ_1;
DR PROSITE: PS01135; FTSZ_2;
KW Cell division; Septation; GTP-binding.
FT NP_BIND 107 115 GTP (POTENTIAL).
SQ SEQUENCE 592 AA; 63503 MW; C32007DADDCD275B CRC64;

alignment_scores:
Quality: 56.50 Length: 40
Ratio: 2.173 Gaps: 2
Percent Similarity: 65.000 Percent Identity: 32.500

alignment_block:
US-09-684-016-48411/rev x FTSZ_BARBA ..

Align seg 1/1 to: FTSZ_BARBA from: 1 to: 592

119 GTGTGATAATAATGAGATAGATACCTTA.....CAGTGAACCT 76
:::|||||:|||||:|||||:|||||:|||||:|||||:
505 LeuTrpGlnArgLeuGlnSerLeuThrTyrArgGlnGluLeuGluP 521
75 TACAGTCAGTGCACCCCTGCATTTGCCCATTAAGTCACAGGTTGTA 26
:::::|||||:|||||:|||||:|||||:|||||:
521 OGluAlaArgLeuGlnProAlaValLysProLeuGlnAsnGlu..... 535

25 AGGCATCTCACATTTCACAT 6
|||||:|||||:|||||:|||||:|||||:
536 .GluserHisIleTyrAsn 541

seq_name: SwissProt_39:GCKR_XENLA
seq_documentation_block:
ID GCKR_XENLA STANDARD: PRT; 619 AA.
AC O91754:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)

```


DE GLUCOKINASE REGULATORY PROTEIN (GLUCOKINASE REGULATOR).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 OC Xenopodinae; Xenopus.
 NX NCBI_TaxID=8335;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=95010134; PubMed=7925465;
 RA Veiga-Da-Cunha M., Delheux M., Watelet N., van Schaftingen E.;
 RT Cloning and expression of a Xenopus liver cDNA encoding a fructose-
 phosphate-insensitive regulatory protein of glucokinase.";
 RL Eur. J. Biochem. 225:43-51(1994).
 CC -1- FUNCTION: INHIBITS GLUCOKINASE BY FORMING AN INACTIVE COMPLEX WITH
 THIS ENZYME.
 CC -1- SIMILARITY: BELONGS TO THE SIS FAMILY. GCKR SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: X80901; CAA56863.1; -
 CC DR InterPro: IPR001741; GCKR.
 CC DR InterPro: IPR001347; SIS.
 CC DR Pfam: PF01380; SIS; 1.
 CC DR ProDom: PD025295; GCKR; 1.
 CC DR PROSITE: PS01272; GCKR; 1.
 SQ SEQUENCE 619 AA; 68738 MW; 41B72C1981D1BA52 CRC64;

alignment_scores: Quality: 56.50 Length: 28
 Ratio: 2.568 Gaps: 1
 Percent Similarity: 78.571 Percent Identity: 39.286

alignment_block:

US-09-684-016-48411 x GCKR_XENLA ..

Align seg 1/1 to: GCKR_XENLA from: 1 to: 619

```

3  GGATTTGTAAT...GTGAGTCCTTACCAACCTGATGACCTAATGCGC 49
   |||||.....:|||||:|||||:|||||:|||||:|||||:|||||
341 G|Y|L|E|E|T|S|P|A|L|A|G|U|C|Y|S|V|A|P|P|O|T|H|Y|R|G|I|N|A|A|S|P|R|P|A|R 357
   |||||:|||||:|||||:|||||:|||||:|||||:|||||
50  CAATGCGAGGCGTGCCTGACTGACTGTAAGTTTCAC 83
   |||||:|||||:|||||:|||||:|||||:|||||:|||||
357 G|A|S|P|V|A|L|R|G|L|P|H|E|I|E|H|T|H|G|L|Y|T|Y|R|H|S 368

```

seq_name: SwissProt_39:OPLA_RAT

seq_documentation_block:

ID OPLA_RAT STANDARD; PRT; 1288 AA.
 AC P97608;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE 15-DEC-1998 (Rel. 37, Last annotation update)
 DE 5-OXOPROLINASE (EC 3.5.2.9) (5-OXO-L-PROLINASE) (PYROGLUTAMASE)
 GN (5-OPAS).
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 NCBI_TaxID=10116;
 RN (1)
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Kidney;
 RX MEDLINE=97113037; PubMed=8943290;
 RA Ye G.-J., Breslow E., Meister A.;

RT "The amino acid 'sequence of rat kidney 5-oxo-L-prolinase determined
 RT by cDNA cloning.";
 RL J. Biol. Chem. 271:32293-32300(1996).
 RN (2)
 RP ERRATUM.
 RA Ye G.-J., Breslow E., Meister A.;
 RL J. Biol. Chem. 272:4646-4646(1997).
 CC -1- FUNCTION: CATALYZES THE CLEAVAGE OF 5-OXO-L-PROLINE TO FORM L-
 CC GLUTAMATE COUPLED TO THE HYDROLYSIS OF ATP TO ADP AND INORGANIC
 CC PHOSPHATE.
 CC -1- CATALYTIC ACTIVITY: ATP + 5-OXO-L-PROLINE + 2 H(2)O = ADP +
 CC PHOSPHATE + L-GLUTAMATE.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- TISSUE SPECIFICITY: WELL EXPRESSED IN TESTIS, KIDNEY AND LIVER.
 CC -1- SIMILARITY: BELONGS TO THE OXOPROLINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: U70825; AAC52955.1; -
 CC DR InterPro: IPR002821; Hydanolase_A.
 CC DR InterPro: IPR003692; Hydanolase_B.
 CC DR Pfam: PF01968; Hydanolase_A; 1.
 CC DR Pfam: PF02538; Hydanolase_B; 1.
 CC DR Hydrolase.
 SQ SEQUENCE 1288 AA; 137746 MW; 4530D0781E10C7AD CRC64;

alignment_scores: Quality: 56.00 Length: 30
 Ratio: 2.800 Gaps: 1
 Percent Similarity: 66.667 Percent Identity: 40.000

alignment_block:

US-09-684-016-48411/rev x OPLA_RAT ..

Align seg 1/1 to: OPLA_RAT from: 1 to: 1288

```

86  CATGTGAACCTTACCAGTACGACGACCT.....CGCATGGCCGCAATT 43
   |||||:|||||:|||||:|||||:|||||:|||||:|||||
215 H|I|S|V|A|L|S|E|R|U|S|E|R|S|E|R|G|I|V|A|L|E|R|P|O|T|H|Y|R|G|I|N|A|A|S|P|R|P|A|R 231
   |||||:|||||:|||||:|||||:|||||:|||||:|||||
42  AGTTCATCAGGCTGTGAAGCATCTCATATTACAAATPAC 3
   |||||:|||||:|||||:|||||:|||||:|||||:|||||
231 G|L|Y|H|I|S|T|H|R|A|C|Y|S|A|L|A|S|P|R|A|L|A|T|Y|L|E|U|T|H|R|P|R|O|T|R 244

```

seq_name: SwissProt_39:H1P1_HUMAN

seq_documentation_block:

ID H1P1_HUMAN STANDARD; PRT; 995 AA.
 AC 000291; O00328;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HUNTINGTIN INTERACTING PROTEIN 1 (HIP-1) (FRAGMENT).
 GN H1P1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Hug A.H.M.M., Nichol K., Osborne L., Scherer S.W., Squitieri F.,
 RA Hayden M.R.;
 RT Genomic organization of the human H1P1 gene and its exclusion as a
 RT candidate gene in a family diagnosed with Huntington disease without
 RT CAG expansion.";
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.

KA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoussis K.
RA Anodaga T. Anodharymme E. C. Tis T. Ver C. Geburt D. C.

03 Bacteria: Proteobacteria: gamma subdivision: E

OC *Escherichia*.

OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=93347969; PubMed=8346018;
 RA Plunkett G., Irl, Burland V.D., Daniels D.L., Blattner F.R.;
 RT "Analysis of the Escherichia coli genome. III. DNA sequence of the
 RT region from 87.2 to 89.2 minutes.";
 RL Nucleic Acids Res. 21:3391-3398(1993).
 RN [2]
 RP SEQUENCE OF 1-5, AND CHARACTERIZATION.
 RX MEDLINE=94291668; PubMed=8020507;
 RA Palmer T., Vasishta A., Whitty P.W., Boxer D.H.;
 RT "Isolation of protein FA, a product of the mob locus required for
 RT molybdenum cofactor biosynthesis in Escherichia coli.";
 RL Eur. J. Biochem. 222:687-692(1994).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=96004466; PubMed=7551035;
 RA Lobbi-Nivol C., Palmer T., Whitty P.W., McNaughton E., Boxer D.H.;
 RT "The mob locus of Escherichia coli K12 required for molybdenum
 RT cofactor biosynthesis is expressed at very low levels.";
 RL Microbiology 141:1663-1671(1995).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).
 RX MEDLINE=20568278; PubMed=10978347;
 RA Late M.W., Temple C.A., Rajagopalan K.V., Schindelin H.;
 RT "The crystal structure of the Escherichia coli MobA protein provides
 RT insight into molybdopterine guanine dinucleotide biosynthesis.";
 RL J. Biol. Chem. 275:40211-40217(2000).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (1.35 ANGSTROMS).
 RX PubMed=11080634;
 RA Stevenson C.E., Sargent F., Buchanan G., Palmer T., Lawson D.M.;
 RT "Crystal structure of the molybdenum cofactor biosynthesis protein
 RT MobA from Escherichia coli at near-atomic resolution.";
 RL Structure 8:1115-1125(2000).
 CC -1- FUNCTION: LINKS A GUANOSINE 5'-PHOSPHATE TO MOLYBDOTERIN (MP)
 CC -1- COFACTOR: REQUIRES MAGNESIUM FOR ACTIVITY.
 CC -1- PATHWAY: MOLYBDENUM COFACTOR BIOSYNTHESIS.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO THE MOBA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L19201; AAB02992.1; -
 DR EMBL: AE000461; AAC76855.1; -
 DR PIR: S40803; S40803.
 DR PDB: 1ESK; 11-NOV-00.
 DR PDB: 1FR9; 10-JAN-01.
 DR PDB: 1FRW; 10-JAN-01.
 DR Ecogene: EGI1829; mobA.
 KW Molybdenum cofactor biosynthesis; GTP-binding; Magnesium;
 KW 3D-structure; Complete proteome.
 SO SEQUENCE 194 AA; 21643 MW; B/9B32D7348DD48 CRC64;

alignment_scores:
 Quality: 52.00 Length: 16
 Ratio: 4.727 Gaps: 0
 Percent Similarity: 68.750 Percent Identity: 50.000

alignment_block:
 US-09-684-016-48411 x MOBA_ECOLI ..

Align seg 1/1 to: MOBA_ECOLI from: 1 to: 194
 5 TATGTGAATGTCGAGATCCCTTACACCGTCGATGACCTAATGCGCAA 52
 :||| ||||| ||| ||||| |||||
 97 PhcysProcsAspThrProtyrIleProProAspPheulAlaIatay 112
 seq_name: SwissProt_39:SUS2_PEA
 seq_documentation_block:
 ID SUS2_PEA STANDARD; PRT; 809 AA.
 AC 024301;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE SUCROSE SYNTHASE 2 (EC 2.4.1.13) (SUCROSE-UDP GLUCOSYLTRANSFERASE 2).
 GN SUS2.
 OS Pisum sativum (Garden pea).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Euphorbia; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
 OX NCBI_TaxID=3888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. FRISSON; TISSUE=seed coat;
 RA Buchner P., Poret M., Rochat C.;
 RT "Cloning and characterization of a cDNA encoding a second sucrose
 RT synthase gene in pea (Pisum sativum L.).";
 RL (in) Plant Gene Register PGR98-105.
 CC -1- FUNCTION: SUCROSE-CLEAVING ENZYME THAT PROVIDES UDP-GLUCOSE AND
 CC FRUCTOSE FOR VARIOUS METABOLIC PATHWAYS.
 CC -1- CATALYTIC ACTIVITY: UDP-GLUCOSE + D-FRUCTOSE = UDP + SUCROSE.
 CC -1- SIMILARITY: TO OTHER PLANTS SUCROSE SYNTHASE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ001071; CAA04512.1; -
 DR InterPro: IPR001286; Glycos_transf_1.
 DR InterPro: IPR000368; Sucrose_synth.
 DR Pfam: PF00534; Glycos_transf_1; 1.
 DR Pfam: PF00862; Sucrose_synth; 1.
 KW Transferase; Glycosyltransferase; Multigene family.
 SO SEQUENCE 809 AA; 92208 MW; 6C5BBF708C37DE75 CRC64;

alignment_scores:
 Quality: 52.00 Length: 57
 Ratio: 2.364 Gaps: 2
 Percent Similarity: 38.596 Percent Identity: 26.316

alignment_block:
 US-09-684-016-48411/rev x SUS2_PEA ..

Align seg 1/1 to: SUS2_PEA from: 1 to: 809

105 GAAGATAGATGATACCTTACATGTGAA..... 79
 ||||| ||||| ||||| ||||| |||||
 457 GluAspStyThrHisPheSerCysGlnPheThrAlaAspLeuIleAla 473
 78 ACTTACGATGATG..... 64
 ||||| ||||| ||||| ||||| |||||
 473 tAsnAsnAlaAspPheIleIleThrSerThrTyrgInGluIleAlaGlyT 490
 63 CAGCCCTGCGATGCGCCATGAGTCATCA 34
 ||||| ||||| ||||| ||||| |||||
 490 hrlYsAsnThrIleGlyGlnTyrgLusSerHisThrAlaPheThrLeuPro 506
 33 GGGTTGTAAGGATTCACAT 13

|||||
507 GlyLeuTYrArgValAlHis 513

seq_name: SwissProt_39:FBP2_STRPU

seq_documentation_block:
ID FBP2_STRPU STANDARD; PRT; 325 AA.
AC P15216;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE FIBROPELLEIN II PRECURSOR (EPIDERMAL GROWTH FACTOR-RELATED PROTEIN 2).
GN EGF2.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90049203; PubMed=2814501;
RA Yang O., Angerer L.M., Angerer R.C.;
RT "Unusual pattern of accumulation of mRNA encoding EGF-related protein
in sea urchin embryos.";
RL Science 246:806-808(1989).
CC -1- SIMILARITY: CONTAINS 4 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: EACH OF THE EGF-LIKE DOMAINS IS VERY SIMILAR TO THE
CC EXOGASTRULA-INDUCING PEPTIDES FROM THE SEA URCHIN ANTHOCIDARIS
CC CRASSISTEIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M29004; AAA30044.1; -.
DR HSSP: P04070; LAUT.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR Pfam: PF00008; EGF; 3.
DR SMART: SM00181; EGF; 4.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2; 4.
KW EGF-like domain; Repeat; Signal.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 325 FIBROPELLEIN II.
FT DOMAIN 48 91 EGF-LIKE 1.
FT DOMAIN 107 152 EGF-LIKE 2.
FT DOMAIN 178 224 EGF-LIKE 3.
FT DOMAIN 252 293 EGF-LIKE 4.
FT DISULFID 52 65 BY SIMILARITY.
FT DISULFID 59 75 BY SIMILARITY.
FT DISULFID 77 90 BY SIMILARITY.
FT DISULFID 111 124 BY SIMILARITY.
FT DISULFID 118 137 BY SIMILARITY.
FT DISULFID 139 151 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.
FT DISULFID 189 209 BY SIMILARITY.
FT DISULFID 211 223 BY SIMILARITY.
FT DISULFID 256 269 BY SIMILARITY.
FT DISULFID 263 278 BY SIMILARITY.
FT DISULFID 280 292 BY SIMILARITY.
SQ SEQUENCE 325 AA; 36898 MW; F09BA805654C7AC CRC64;

alignment_scores:
Quality: 51.50 Length: 14
Ratio: 4.682 Gaps: 1
Percent Similarity: 78.571 Percent Identity: 64.286

alignment_block:
US-09-684-016-48411 x FBP2_STRPU ..

Align seg 1/1 to: FBP2_STRPU from: 1 to: 325

5 TATGTGAATGTGAGATGCTTACAAAC.....CCTGAT 37
|||||
74 TyrCysLysCysGluMetProPheArgValGlyIleProAsp 87

seq_name: SwissProt_39:YDDB_SCHPO

seq_documentation_block:
ID YDDB_SCHPO STANDARD; PRT; 859 AA.
AC O10362;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOETHERICAL 94.9 KDA PROTEIN C22E12.11C IN CHROMOSOME I.
GN SPAC22E12.11C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Devlin K., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 PHD-TYPE ZINC FINGER.
CC -1- SIMILARITY: TO YEAST YKR029C AND YJL105W.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Z70043; CA93898.1; -.
DR InterPro: IPR001965; PHD.
DR InterPro: IPR001214; SET.
DR Pfam: PF00628; PHD; 1.
DR Pfam: PF00856; SET; 1.
DR SMART: SM00249; PHD; 1.
DR SMART: SM00317; SET; 1.
DR PROSITE: PS50280; SET; 1.
KW Hypothetical protein; Zinc-finger.
SQ SEQUENCE 859 AA; 94886 MW; B4865BF40FD2C5D8 CRC64;

alignment_scores:
Quality: 51.50 Length: 22
Ratio: 3.679 Gaps: 1
Percent Similarity: 63.636 Percent Identity: 40.909

alignment_block:

US-09-684-016-48411 x YDDB_SCHPO ..

Align seg 1/1 to: YDDB_SCHPO from: 1 to: 859

11 AAATGTGATGCTTACACCCGTGATAC...CTAATGGCCGATGGCA 57
|||||
5 ArgCysValCysProPheGluAspAspGlyPheThrIleGlnCysG1 21

58 GGGCTGCACCTGACTGG 73
|||||
21 uSerCysGluValIrrp 26

seq_name: SwissProt_39:HEX9_ADECC

```

seq_documentation_block:
ID  HEX9_ADECC  STANDARD:  PRT:  103 AA.
AC  Q65944;
DT  01-NOV-1997 (Rel. 35, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DT  15-DEC-1998 (Rel. 37, Last annotation update)
DE  HEXON-ASSOCIATED PROTEIN (PROTEIN IX).
GN  PIX.
OS  Canine adenovirus type 1 (strain CL1), and
OS  Canine adenovirus type 1 (strain RI261).
OC  Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX  NCBI_TaxID=69150, 69151;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CL1;
RA  Campbell J.B., Zhao Y.;
RL  Submitted (AUG-1996) to the EMBL/Genbank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=RI261;
RA  MEDLINE=9129661;
RA  Morrison M.D., Onions D.E., Nicolson L.;
RT  *Complete DNA sequence of canine adenovirus type 1.;
RL  J. Gen. Virol. 78:873-878(1997).
CC  -!- FUNCTION: THIS PROTEIN IS A STRUCTURAL COMPONENT OF THE VIRION. IT
CC  MAY HAVE AN ADDITIONAL ROLE DURING ADENOVIRUS MULTIPLICATION.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: U55001; AAB05432.1; -
DR  EMBL: Y07760; CAA69055.1; -
KW  Hexon-associated protein.
SQ  SEQUENCE 103 AA; 11303 MW; 4C14502AD6F5A8C1 CRC64;

alignment_scores:
Quality: 51.00 Length: 24
Ratio: 3.188 Gaps: 0
Percent Similarity: 66.667 Percent Identity: 50.000

alignment_block:
US-09-684-016-48411 x HEX9_ADECC ..
Align seg 1/1 to: HEX9_ADECC from: 1 to: 103

3 GGTATTGTAATGTGAGATGCCTTACACCCGTGATGACCTAATGCGCAA 52
|||||
7 GYLLVLAASnThrcysPheLeuThrArgIleProSerIrrPaaGI 23
|||||
53 TCGAGCGCTGCACGTGCT 74
|||||
23 YALAArGlnAsnValThrCly 30
|||||

seq_name: SwissProt_39:DMSB_ECOLI
seq_documentation_block:
ID  DMSB_ECOLI  STANDARD:  PRT:  204 AA.
AC  P18776; P77745;
DT  01-NOV-1990 (Rel. 16, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DT  20-AUG-2001 (Rel. 40, Last annotation update)
DE  ANAEROBIC DIMETHYL SULFOXIDE REDUCTASE CHAIN B (EC 1.-.-.-) (DMSO
DE  REDUCTASE IRON-SULFUR SUBUNIT).
GN  DMSB OR B0895.
OS  Escherichia coli.
OC  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC  Escherichia.

```

```

OX  NCBI_TaxID=562;
RN  [1]
RP  SEQUENCE FROM N.A., AND SEQUENCE OF 1-9.
RC  STRAIN=K12 / C600;
RC  MEDLINE=89096500; PubMed=3062312;
RA  Bilous P.T., Cole S.T., Anderson W.F., Weiner J.H.;
RT  *Nucleotide sequence of the dmsABC operon encoding the anaerobic
RT  dimethylsulphoxide reductase of Escherichia coli.;
RL  Mol. Microbiol. 2:785-795(1988).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=K12 / MG1655;
RC  MEDLINE=97426617; PubMed=9278503;
RA  Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA  Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA  Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA  Mau B., Shao Y.;
RT  *The complete genome sequence of Escherichia coli K-12.;
RL  Science 277:1453-1474(1997).
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN=K12;
RC  MEDLINE=97061202; PubMed=8905232;
RA  Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA  Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA  Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA  Mori H., Motomura K., Nakamura Y., Nashimoto H., Nishio Y., Saito N.,
RA  Sano M., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA  Yano M., Horuchi T.;
RT  *A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT  corresponding to the 12.7-28.0 min region on the linkage map.;
RL  DNA Res. 3:137-155(1996).
RN  [4]
RP  EPR SPECTROSCOPY OF IRON-SULFUR CLUSTERS.
RC  MEDLINE=91070067; PubMed=2174699;
RA  Cammack R., Weiner J.H.;
RT  *Electron paramagnetic resonance spectroscopic characterization of
RT  dimethyl sulfoxide reductase of Escherichia coli.;
RL  Biochemistry 29:8410-8416(1990).
RN  [5]
RP  MUTAGENESIS.
RC  MEDLINE=91355180; PubMed=1653010;
RA  Rothery R.A., Weiner J.H.;
RT  *Alteration of the iron-sulfur cluster composition of Escherichia
RT  coli dimethyl sulfoxide reductase by site-directed mutagenesis.;
RL  Biochemistry 30:8296-8305(1991).
CC  -!- FUNCTION: TERMINAL REDUCTASE DURING ANAEROBIC GROWTH ON
CC  VARIOUS SULFOXIDE AND N-OXIDE COMPOUNDS.
CC  -!- CORFACTOR: IRON-SULFUR.
CC  -!- SUBUNIT: THE COMPLEX CONSIST OF THREE SUBUNITS: DMSA, THE
CC  REDUCTASE; DMSB, AN ELECTRON TRANSFER PROTEIN, AND DMSB, A
CC  MEMBRANE ANCHOR PROTEIN.
CC  -!- SIMILARITY: THE IRON-SULFUR CENTERS ARE SIMILAR TO THOSE OF
CC  'BACTERIAL-TYPE' 4FE-4S FERREDOXINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: J03412; AAC83844.1; -
DR  EMBL: AE000191; AAA33981.1; -
DR  EMBL: D90727; BAA35627.1; -
DR  PIR: S03786; S03786.
DR  HSP: P00195; ICLF.
DR  EcoGene: EG10233; dmsB.
DR  InterPro: IPR001450; 4FE4S_ferriXin.
DR  Pfam: PF00037; fer4; 1.
DR  PROSITE: PS00198; 4FE4S_FERREDOXIN; 1.
KW  Oxidoreductase; Electron transport; 4fe-4s; Iron-sulfur;

```

CC SUBFAMILY.

RL J. Virol. 66:1665

```

RP PHOSPHORYLATION.
RX MEDLINE-92188523; PubMed-1532104;
RA Krajeski P., Mold W.S.M.;
RT "The adenovirus E3-14.5K protein which is required for prevention of
RT TNF cytotoxicity and for down-regulation of the EGF receptor contains
RT phosphoserine."
RL Virology 187:492-498(1992).
RN [5]
RP O-GLYCOSYLATION.
RX MEDLINE-92263760; PubMed-1533979;
RA Krajeski P., Tollefson A.E., Mold W.S.M.;
RT "The E3-14.5K integral membrane protein of adenovirus that is
RT required for down-regulation of the EGF receptor and for prevention
RT of TNF cytotoxicity is O-glycosylated but not N-glycosylated."
RL Virology 188:570-579(1992).
RN [6]
RP COMPLETE GENOME.
RX MEDLINE-92087470; PubMed-1727603;
RA Chroboczek J., Biebler F., Jacrot B.;
RT "The sequence of the genome of adenovirus type 5 and its comparison
RT with the genome of adenovirus type 2."
RL Virology 186:280-285(1992).
CC -1- FUNCTION: DOWN-REGULATES THE EGF RECEPTOR AND PREVENTS CYTOLYSIS
CC BY TNF.
CC -1- PTM: PHOSPHORYLATED ON SERINE; O-GLYCOSYLATED, BUT NOT
CC N-GLYCOSYLATED.
CC -1- SIMILARITY: TO EQUIVALENT PROTEIN IN OTHER ADENOVIRUSES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M73260; -; NOT_ANNOTATED_CDS.
DR EMBL; X03002; CA26786.1; -.
KW Early protein; Transmembrane; Glycoprotein; Phosphorylation; Signal.
FT SIGNAL 1 18
FT CHAIN 19 132 EARLY E3B 14.6 KDA PROTEIN.
FT TRANSMEM 51 75 POTENTIAL.
FT SEQUENCE 132 AA; 14750 MW; 524690C4AD9B9A74 CRC64;

alignment_scores:
Quality: 50.50 Length: 37
Ratio: 2.196 Gaps: 1
Percent Similarity: 62.162 Percent Identity: 29.730

alignment_block:
US-09-684-016-48411/rev x E315_ADE05 ..
Align seg 1/1 to: E315_ADE05 from: 1 to: 132

107 TTGAAGATAGATATACAC.....TTACATGTGAAGAACTTACACAGTCA 67
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MetLysPheThrValThrPheLeuLeuIleIleCysThrLeuSerAlaPhe 17
66 GGAGCGCCCTGGCATTCGGCCATTAAGTCATCAGGTTGTAAGGCATCTC 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
17 eCySerProThrSerLysPheGlnArgHisIleSerCysArgPheThrA 34
16 ACATTTCACAT 6
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
34 gLleTirPasn 37

seq_name: SwissProt_39:SPC2_CHICK
seq_documentation_block:
ID SPC2_CHICK STANDARD; PRT; 180 AA.
AC P28687;
DT 01-DEC-1992 (Rel. 24, Created)

```

```

DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE MICROSOVAL SIGNAL PEPTIDASE 23 KDA SUBUNIT (EC 3.4.-.-) (SPC22/23)
DE (GP23).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Oviduct;
RX MEDLINE-92189580; PubMed-1546959;
RA Newsome A.L., McLean J.W., Lively M.O.;
RT "Molecular cloning of a cDNA encoding the glycoprotein of hen oviduct
RT microosomal signal peptidase."
RL Biochem. J. 282:447-452(1992).
CC -1- FUNCTION: MICROSOVAL SIGNAL PEPTIDASE IS A MEMBRANE-BOUND
CC ENDOPEPTIDASE THAT REMOVES SIGNAL PEPTIDES FROM NASCENT
CC PROTEINS AS THEY ARE TRANSLOCATED INTO THE LUMEN OF THE
CC ENDOPLASMIC RETICULUM.
CC -1- SUBUNIT: HEN OVIDUCT SIGNAL PEPTIDASE ACTIVITY IS ASSOCIATED
CC WITH AN APPARENT COMPLEX OF ONLY TWO PROTEINS, WITH ESTIMATED
CC MOLECULAR MASSES OF 19 AND 23 KDA.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MICROSOVAL.
CC -1- SIMILARITY: TO OTHER VERTEBRATES SPC22/23 AND TO YEAST SPC3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X60795; CAA43208.1; -.
DR PIR; S15551; S15551.
DR PIR; S22412; S22412.
KW Glycoprotein; Hydrolase; Microsome; Endoplasmic reticulum;
KW Transmembrane; Protease; Signal-anchor.
FT DOMAIN 1 10 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 11 32 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT DOMAIN 33 180 (POTENTIAL).
FT FT 144 164 LUMENAL (POTENTIAL).
FT CARBOHYD 141 141 HYDROPHOBIC.
FT SEQUENCE 180 AA; 20230 MW; A5572439EB8210AA CRC64;

alignment_scores:
Quality: 50.00 Length: 28
Ratio: 2.500 Gaps: 0
Percent Similarity: 71.429 Percent Identity: 39.266

alignment_block:
US-09-684-016-48411 x SPC2_CHICK ..
Align seg 1/1 to: SPC2_CHICK from: 1 to: 180

38 GACCTATATGGCGAATGCGAGGCGTCACTGATGTAAGTTCACATCT 87
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
69 AspieThrAlaAspLeuGlnSerIlePheAspTirPasnValIysGlnIle 85
88 AAGGTGATACCTTATCTTCAATTATTCACACA 121
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
85 upheLeuTyrLeuSerAlaGlnTyrSerThrIys 96

seq_name: SwissProt_39:NGFR_HUMAN
seq_documentation_block:
ID NGFR_HUMAN STANDARD; PRT; 427 AA.
AC P08138;
DT 01-AUG-1988 (Rel. 08, Created)

```



```

FT DISULFID 89 107 BY SIMILARITY.
FT DISULFID 109 122 BY SIMILARITY.
FT DISULFID 125 138 BY SIMILARITY.
FT DISULFID 128 146 BY SIMILARITY.
FT DISULFID 149 164 BY SIMILARITY.
FT DISULFID 167 180 BY SIMILARITY.
FT DISULFID 170 188 BY SIMILARITY.
FT DOMAIN 197 248 SER/THR-RICH.
SQ CARBOHYD 60 60 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 427 AA; 45183 MW; B09FA143FB3D625B CRC64;

alignment_scores:
    Quality: 50.00 Length: 19
    Ratio: 4.167 Gaps: 0
Percent Similarity: 63.158 Percent Identity: 36.842

alignment_block:
US-09-684-016-48411 x NGFR_HUMAN ..

Align seg 1/1 to: NGFR_HUMAN from: 1 to: 427

      8 TGTATATGTGAGATGCTTACACACCTGATGACCTAATGCGAATGCGA 57
      |||:||||| ||| |||:|||||
107 CysATgCysAlaTyrGlyTyrTyrGlnAspGluThrThrGlyArgCysGln 123
      58 GGGCTGTC 64
      ||:||||
123 ualacYcs 125

seq_name: SwissProt_39:PHF1_MOUSE

seq_documentation_block:
ID PHF1_MOUSE STANDARD; PRT; 559 AA.
AC G932188: 054808:
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PHD FINGER PROTEIN 1 (PHF1 PROTEIN) (T-COMPLEX TESTIS-EXPRESSED 3)
DE (POLYCOMBLIKE 1) (MPCCL).
GN PHF1 OR TCTEX3 OR TCTEX-3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NIH SWISS;
RA Howard T.L., Ingemann A.R., Hollenberg S.M.;
RT "The Basic helix-loop-helix protein Th1 interacts with polycomb-group
RT proteins.";
RL Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=99018220; PubMed=9799836;
RA Kawakami S., Mitsunaga K., Kikuti Y.Y., Ando A., Inoko H.,
RA Yamamura K., Abe K.;
RT "Tctex3, related to Drosophila polycomblike, is expressed in male germ
RT cells and mapped to the mouse t-complex.";
RL Mamm. Genome 9:874-880(1998).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -!- TISSUE SPECIFICITY: TESTIS-SPECIFIC.
CC -!- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
-----

```

DR EMBL; U81490: AAD00518.1; -
 DR EMBL; AB011550: BAA25074.1; -
 DR MGI; MGI:98647; Tctex3.
 DR InterPro: IPR001965; PHD.
 DR InterPro: IPR002999; Tudor.
 DR Pfam: PF00628; PHD; 2.
 DR SMART; SM00249; PHD; 2.
 DR SMART; SM00333; Tudor; 1.
 KW Nuclear protein; zinc-finger; Repeat.
 FT ZN_FING 89 142 PHD-type 1.
 FT ZN_FING 188 237 PHD-type 2.
 FT CONFLICT 249 249 L -> E (IN REF. 1).
 FT CONFLICT 344 344 G -> S (IN REF. 2).
 FT CONFLICT 377 377 R -> S (IN REF. 2).
 FT CONFLICT 555 555 G -> R (IN REF. 2).
 FT CONFLICT 558 559 IF -> HLPDLSLLLPSPPTHHFHALDL (IN REF. 2).
 SO SEQUENCE 559 AA; 61139 MW; 6D2E5F53D6164C2 CRC64;

alignment_scores:
 Quality: 50.00 Length: 23
 Ratio: 3.571 Gaps: 0
 Percent Similarity: 60.870 Percent Identity: 39.130

alignment_block:

US-09-684-016-48411 x PHF1_MOUSE ..

Align seg 1/1 to: PHF1_MOUSE from: 1 to: 559

5 TATTGTAAATGTGATGCTTACACCCGTGACCTAATGGCGCATG 54
 ||||| ||| ||| :::: ||| ||| |||||
 188 TYRCYSTYRCYSGLYGTYRPGYGLUTRPSNLEULYSMETLEUINC 204

55 CGAGGCGCTGCACCTGACTGG 73
 |:::||||| ::|||
 204 SARGSERCYSLGGLNTRP 210

seq_name: SwissProt_39:PHF1_HUMAN

seq_documentation_block:

ID PHF1_HUMAN STANDARD; PRT; 567 AA.
 AC 043189; 060929;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 DE PHD FINGER PROTEIN 1 (PHF1 PROTEIN).
 GN PHF1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Placenta;
 RX MEDLINE=98207256; PubMed=9545646;
 RA Coulson M., Robert S., Eyre H.J., Saint R.;
 RT "The identification and localization of a human gene with sequence
 RT similarity to Polycomblike of Drosophila melanogaster.";
 RL Genomics 48:381-383(1998).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RA Wang J.H., Du G.W., Zhou Y., Yuan J.G., Qiang B.O.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RA Mashreghi-Mohammadi M.;
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: 1/PHF1 AND 2/PHF2 (SHOWN HERE);
 CC ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN HEART, SKELETAL MUSCLE, AND
 CC PANCREAS, LOWER LEVELS IN BRAIN, PLACENTA, LUNG, LIVER, AND

CC KIDNEY.
 CC -1- SIMILARITY: CONTAINS 2 PHD-TYPE ZINC FINGERS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.sib.ch).

CC EMBL; AF029678: AAC52062.1; -
 CC EMBL; AF052205: AAC13273.1; -
 CC EMBL; AL021366: CAI6158.1; -
 CC EMBL; AL021366: CAI6159.1; -
 CC MIM: 602881; -

DR InterPro: IPR001965; PHD.
 DR InterPro: IPR002999; Tudor.

DR Pfam: PF00628; PHD; 2.
 DR SMART; SM00249; PHD; 2.

KW Nuclear protein; zinc-finger; Repeat; Alternative splicing.

FT ZN_FING 89 142 PHD-type 1.
 FT ZN_FING 188 237 PHD-type 2.

FT VARSPLIC 350 457

FT SPPSGCGGCGVSRPLGRRRPPPEPLRRRKGVEELGPP
 SAVNRQPEQDERAHLOALQASVSPSPSPOSTOGSS
 GYNEPRTDRCILPSPDIRMFASFHS -> RAGPWGRLS
 PGEAPEAGARAPDEEAGEGSCGAGATLSASASAPGAEGA
 GSSAEGTAAPSGCLLPSTLLPAPOGPLGTVDPOTHPMWF
 TLVSPTSLSKVPPR (IN ISOFORM 1).
 FT VARSPLIC 458 567 MISSING (IN ISOFORM 1).
 SO SEQUENCE 567 AA; 62077 MW; 3492A94ADP959579 CRC64;

alignment_scores:

Quality: 50.00 Length: 23
 Ratio: 3.571 Gaps: 0
 Percent Similarity: 60.870 Percent Identity: 39.130

alignment_block:

US-09-684-016-48411 x PHF1_HUMAN ..

Align seg 1/1 to: PHF1_HUMAN from: 1 to: 567

5 TATTGTAAATGTGATGCTTACACCCGTGACCTAATGGCGCATG 54
 ||||| ||| ||| :::: ||| ||| |||||
 188 TYRCYSTYRCYSGLYGTYRPGYGLUTRPSNLEULYSMETLEUINC 204

55 CGAGGCGCTGCACCTGACTGG 73
 |:::||||| ::|||
 204 SARGSERCYSLGGLNTRP 210

seq_name: SwissProt_39:BS4_HUMAN

seq_documentation_block:

ID BS4_HUMAN STANDARD; PRT; 601 AA.
 AC Q9Y5A7; 095422;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE 20-AUG-2001 (Rel. 40, Last annotation update)
 DE BS4 PROTEIN (NY-REN-18 ANTIGEN).
 GN NYREN18.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99438124; PubMed=10508479;
 RA Scanlan M.J., Jordan J.D., Williamson B., Stockert E., Bander N.H.,
 RA Jorgensen V., Gure A.O., Jager D., Jager E., Knuth A., Chen Y.-T.,
 RA Old L.J.;
 RT "Antigens recognized by autologous antibody in patients with renal-

RT cell carcinoma.";
 RL Int. J. Cancer 83:456-464(1999).
 RN [2]
 RP SEQUENCE OF 1-365 FROM N.A.
 RA Rump A., Rosenthal A., Drescher B., Weber J., Schatevov R.,
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 3 UBA DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF150509; AAD42865.1; -
 DR EMBL: AF108083; AAC82474.1; -
 DR InterPro: IPR000449; UBA.
 DR Pfam: PF00627; UBA; 2.
 DR SMART: SM00165; UBA; 1.
 KW Repeat.
 FT DOMAIN 374 414 UBA 1.
 FT DOMAIN 424 456 UBA 2.
 FT DOMAIN 475 515 UBA 3.
 FT CONFLICT 75 75
 FT D -> DNYRTGATIEVPLPRKDKIAKQYSDRLCCEN
 FT CONFLICT 365 365 EVEKVEIERCKAIENGTD (IN REF. 2).
 FT R -> K (IN REF. 2).
 SQ SEQUENCE 601 AA; 69272 MM; D5A6ECAE0B3A75B CRC64;

alignment_scores:

Quality: 50.00 Length: 46
 Ratio: 2.174 Gaps: 2
 Percent Similarity: 50.000 Percent Identity: 36.957

alignment_block:

US-09-684-016-48411/rev x BS4_HUMAN ..

Align seg 1/1 to: BS4_HUMAN from: 1 to: 601

```

113 AATAATGAGATAGATACCTTACATGTGAACCTTACCAGTCAGTG 64
      ::::::::::::::|::|::|::|::|::|::|::|::|::|
368 SerSerLeuLysSerTyrIleLeuIleHisGlnLysTyrPheIleCysCy 384
      |||||
63 CAGC.....CCTGCATTTGGCCCATTAAGTCATCAGGGTGT. 27
      |||||
384 sSerTrpIleuLeuProAlaGlyHisArgLeuGlyLeuArgAlaCysA 401
      |||||
26 .....AAGCATCTCCACATTTACAAT 6
      ||::|::|::|::|::|::|::|
401 spGIyAsnValAspHisAlaIleThrHisIleThrAsn 413

```

seq_name: SwissProt_39:SUSY_BETVU

seq_documentation_block:

ID SUSY_BETVU STANDARD; PRT; 766 AA.
 AC 042652;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE SUCROSE SYNTHASE (EC 2.4.1.13) (SUCROSE-UDP GLUCOSYLTRANSFERASE)
 DE (FRAGMENT).
 GN SSI.
 OS Beta vulgaris (Sugar beet).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Caryophyllidae; Caryophyllales; Chenopodiaceae; Beta.
 OC NCBI_TaxID=3555;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Tap root;

RX MEDLINE=96270366; PubMed=8639746;
 RA Hesse H., Willmitzer L.;
 RT "Expression analysis of a sucrose synthase gene from sugar beet (Beta
 RL vulgaris L.)";
 RL Plant Mol. Biol. 30:863-872(1996).
 CC -1- FUNCTION: SUCROSE-CLEAVING ENZYME THAT PROVIDES UDP-GLUCOSE AND
 CC FRUCTOSE FOR VARIOUS METABOLIC PATHWAYS.
 CC CATALYTIC ACTIVITY: UDP-GLUCOSE + D-FRUCTOSE = UDP + SUCROSE.
 CC -1- TISSUE SPECIFICITY: EXPRESSED MOST PREDOMINANTLY IN TAP ROOT.
 CC -1- SIMILARITY: TO OTHER PLANTS SUCROSE SYNTHASE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X81974; CAA57499.1; -
 DR InterPro: IPR001296; Glycos_transf_1.
 DR InterPro: IPR000368; Sucrose_synth.
 DR Pfam: PF00534; Glycos_transf_1; 1.
 DR Pfam: PF00862; Sucrose_synth; 1.
 DR Transferase; Glycosyltransferase.
 FT NON_TER 1
 FT 1
 SQ SEQUENCE 766 AA; 87240 MM; 4078033ECD8EC4A82 CRC64;

alignment_scores:

Quality: 50.00 Length: 57
 Ratio: 2.273 Gaps: 2
 Percent Similarity: 38.596 Percent Identity: 26.316

alignment_block:

US-09-684-016-48411/rev x SUSY_BETVU ..

Align seg 1/1 to: SUSY_BETVU from: 1 to: 766

```

105 GAGATAGATACCTTACATGTGAA..... 79
      |||||
399 GluAspLysTyrHisPheSerCysGlnPheSerAlaAspLeuMetAlaMe 415
      |||||
78 .....ACTTACCAGTCAGG..... 64
      |||||
415 TasnHisAlaAspPheIleIleThrSerThrTyrGlnIuIleAlaGlyT 432
      |||||
63 .....CAGCCCTGCATTTGCCCATTAAGTCATCA 34
      |||||
432 hrLysAsnThrValGlyGlnTyrGluSerHisLysAlaPheThrPhePro 448
      |||||
33 GGGTTGTAAAGCATCTCACAT 13
      |||||
449 GlyLeuTyrArgValValHis 455

```

seq_name: SwissProt_39:BAR3_CHITE

seq_documentation_block:

ID BAR3_CHITE STANDARD; PRT; 1700 AA.
 AC 003376;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last annotation update)
 DE BALBIANT RING PROTEIN 3 PRECURSOR.
 GN BR3.
 OS Chironomus tentans (Midge).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
 OC Chironomidae; Chironomidae; Chironominae; Chironomus.
 OC NCBI_TaxID=7153;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Salivary gland;


```

OC Pseudomonas.
OX NCBI_TaxID=43306;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SC510;
RX MEDLINE=91008976; PubMed=2211521;
RA Crouzet J., Cameron B., Cauchois L., Rigault S., Rouyez M.-C.,
RT Blanche F., Thibaut D., Debussche L.;
RT "Genetic and sequence analysis of an 8.7-kilobase Pseudomonas
RT denitrificans fragment carrying eight genes involved in
RT transformation of precorrin-2 to cobyrinic acid.";
RL J. Bacteriol. 172:5980-5990(1990).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=94042919; PubMed=8226690;
RA Debussche L., Thibaut D., Cameron B., Crouzet J., Blanche F.;
RT "Biosynthesis of the corrin macrocycle of coenzyme B12 in Pseudomonas
RT denitrificans.";
RL J. Bacteriol. 175:7430-7440(1993).
CC -1- FUNCTION: METHYLTRANSFERASE THAT CATALYZES THE METHYLATION OF
CC C-17 IN PRECORRIN-3b TO FORM PRECORRIN-4.
CC -1- CATALYTIC ACTIVITY: S-ADENOSYL-L-METHIONINE + PRECORRIN-3b = S-
CC ADENOSYL-L-HOMOCYSTEINE + PRECORRIN 4.
CC -1- PATHWAY: COBALAMIN BIOSYNTHESIS.
CC -1- SIMILARITY: TO S-TYPIFORMIUM CBH; ALSO, LOW, TO OTHER
CC METHYLASES INVOLVED IN COBALAMIN BIOSYNTHESIS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: M59301; AAA25798.1; -
CC PIR: E36145; E36145.
CC InterPro: IPR000878; Corrin_porph_methyltrnf.
CC Pfam: PF00590; TP_methylase; 1.
CC Cobalamin biosynthesis; Porphyrin biosynthesis; Transferrase;
CC Methyltransferase.
CC KW SEQUENCE 254 AA; 27105 MW; A30B027F4AFBE23F CRC64;
SO

```

```

alignment_scores:
  Quality: 49.50      Length: 34
  Ratio: 2.475       Gaps: 1
  Percent Similarity: 58.824   Percent Identity: 38.235

```

alignment_block:

us-09-684-016-48411/rev x COBJ_PSEDE ..

Align seg 1/1 to: COBJ_PSEDE from: 1 to: 254

```

116 TGAATAAATGAGATGATACCTTACATGGAACCTTACAGTCA 67
|||||
103 TrrysseValGluleuValIleThrProGlyValIleAlaMetIleAla 119
66 GTGACGCCCTGCGATT...GCCCATTAAGTCATCAGGGTTGTAGAGCAT 20
::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
119 aValAlaAlaArgIleGlyAlaProLeuGlyHisAspPheCysAlaIle 136
19 CT 18
||
136 er 136

```

seq_name: SwissProt_39:CAIF_HUMAN

seq_documentation_block:

```

ID CAIF_HUMAN STANDARD; PRT; 1603 AA.
AC 007092;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)

```

```

DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE COLLAGEN ALPHA 1(XVI) CHAIN PRECURSOR.
GN COL1A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92335339; PubMed=1631157;
RA Pan T.C., Zhang R.Z., Mattei M.-G., Timpl R., Chu M.-L.;
RT "Cloning and chromosomal location of human alpha 1(XVI) collagen.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:6565-6569(1992).
RN [2]
RP SEQUENCE OF 418-1603 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=93203161; PubMed=1284248;
RA Yamaguchi N., Kimura S., McBride O.W., Hori H., Yamada Y.,
RA Kanamori T., Yamakoshi H., Nagai Y.;
RT "Molecular cloning and partial characterization of a novel collagen
RT chain, alpha 1(XVI), consisting of repetitive collagenous domains and
RL cysteine-containing non-collagenous segments.";
CC J. Biochem. 112:856-863(1992).
CC -1- FUNCTION: THE NUMEROUS INTERRUPTIONS IN THE TRIPLE HELIX MAY MAKE
CC THIS MOLECULE EITHER ELASTIC OR FLEXIBLE.
CC -1- TISSUE SPECIFICITY: IN THE PLACENTA, WHERE IT IS FOUND IN THE
CC AMNION, A MEMBRANOUS TISSUE LINING THE AMNIOTIC CAVITY. WITHIN THE
CC AMNION, IT IS FOUND IN AN ACCELULAR, RELATIVELY DENSE LAYER OF A
CC COMPLEX NETWORK OF RETICULAR FIBERS. ALSO LOCATED TO A FIBROBLAST
CC LAYER BENEATH THIS DENSE LAYER. EXISTS IN TISSUES IN ASSOCIATION
CC WITH OTHER TYPES OF COLLAGEN.
CC -1- DEVELOPMENTAL STAGE: TRANSIENTLY ELEVATED EXPRESSION DURING
CC GESTATION, AND DECREASE AT TERM.
CC -1- DOMAIN: THIS SEQUENCE DEFINES EIGHTEEN DIFFERENT DOMAINS, NINE
CC TRIPLE-HELICAL DOMAINS (COL9 TO COL1) AND TEN NONTRIPLE-HELICAL
CC DOMAINS (NC10 TO NC1).
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- SIMILARITY: BELONGS TO THE FAMILY OF FIBRIL-ASSOCIATED COLLAGENS
CC WITH INTERRUPTED HELICES (FACIT).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: M92642; AAA58427.1; -
CC EMBL: S57132; AAB25797.1; -
CC PIR: S23810; S23810.
CC MIM: 120326; -
CC InterPro: IPR000087; Collagen.
CC InterPro: IPR003129; TSPN.
CC Pfam: PF01391; Collagen; 16.
CC Pfam: PF02210; TSPN; 1.
CC SMART: SM00210; TSPN; 1.
CC KW Extracellular matrix; Connective tissue; Collagen; Hydroxylation;
CC Repeat; Signal.
CC FT SIGNAL 1 21
CC FT CHAIN 22 1603
CC FT DOMAIN 22 374
CC FT DOMAIN 375 505
CC FT DOMAIN 506 520
CC FT DOMAIN 521 554
CC FT DOMAIN 555 571
CC FT DOMAIN 572 630
CC FT DOMAIN 631 651
CC FT DOMAIN 652 722

```

FT	DOMAIN	723	737	WITH 1 IMPERFECTIONS.
FT	DOMAIN	738	875	TRIPE-HELICAL REGION 6 (NC6).
FT	DOMAIN			TRIPE-HELICAL REGION 5 (COL5)
FT	DOMAIN			WITH 3 IMPERFECTIONS.
FT	DOMAIN	876	886	NONHELICAL REGION 5 (NC5).
FT	DOMAIN	887	938	TRIPE-HELICAL REGION 4 (COL4)
FT	DOMAIN			WITH 2 IMPERFECTIONS.
FT	DOMAIN	939	972	NONHELICAL REGION 4 (NC4).
FT	DOMAIN	973	987	TRIPE-HELICAL REGION 3 (COL3)
FT	DOMAIN	988	1010	NONHELICAL REGION 3 (NC3).
FT	DOMAIN	1011	1432	TRIPE-HELICAL REGION 2 (COL2)
FT	DOMAIN			WITH 2 IMPERFECTIONS.
FT	DOMAIN	1433	1471	NONHELICAL REGION 2 (NC2).
FT	DOMAIN	1472	1577	TRIPE-HELICAL REGION 1 (COL1)
FT	DOMAIN			WITH 2 IMPERFECTIONS.
FT	DOMAIN	1578	1603	NONHELICAL REGION 1 (NC1).
FT	CONFLICT	418	420	ROA -> GGR (IN REF. 2).
FT	CONFLICT	537	537	R -> P (IN REF. 2).
FT	CONFLICT	1160	1160	T -> P (IN REF. 2).
FT	CONFLICT	1163	1163	T -> P (IN REF. 2).
FT	CONFLICT	1165	1165	S -> P (IN REF. 2).
SO	SEQUENCE	1603 AA:	157692 MW:	E27D9A1D4E58BA37 CRC64;

RT for all MIT subfam

SEQUENCE FROM N.A., AND SEQUE

RP SEQUENCE FROM N.A., AND SEQUENCE OF 35-50 AND 428-437.
BC
EUCSIE-Quarry

```

RX MEDLINE-90205829; PubMed-1690843;
RA Lang L.-F., Chamow S.M., Dean J.:
RT "Oocyte-specific expression of mouse Zp-2: developmental regulation
RT of the zona pellucida genes."
RL Mol. Cell. Biol. 10:1507-1515(1990).
CC -1- FUNCTION: ZP2 FORMS WITH ZP1 AND ZP3 THE ZONA PELLUCIDA, IN
CC WHICH ZP2 AND ZP3 COMPLEX INTO COPOLYMERS CROSS-LINKED BY ZP1.
CC ZP2 ACTS AS A SECONDARY SPERM RECEPTOR.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. EXTRACELLULAR
CC MATRIX.
CC -1- TISSUE SPECIFICITY: OOCYTES.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING THE 2-WEEK GROWTH PHASE OF
CC OOGENESIS, PRIOR TO OVULATION.
CC -1- PTM: PROTEOLYTICALLY CLEAVED AFTER FERTILIZATION, AND THIS
CC MODIFICATION, ALONG WITH PRESUMED CHANGES IN ZP3 MAY PLAY AN
CC IMPORTANT ROLE IN THE POSTFERTILIZATION BLOCK TO POLYSPERMY.
CC -1- PTM: SULFATED GLYCOPROTEIN WITH O-LINKED OLIGOSACCHARIDES
CC (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 ZP DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M34148; AAA0586.1; -.
DR PIR; A34782; A34782.
DR MGD; MGI:99214; ZP2.
DR InterPro; IPR001507; zona_pellucida.
DR Pfam; PF00100; zona_pellucida.
DR PRINTS; PR00023; ZPELLUCIDA.
DR SMART; SM00241; ZP. 1.
DR PROSITE; PS00682; ZP_DOMAIN: 1.
KW Glycoprotein; Signal; Sulfation; Sperm; Receptor; Transmembrane;
KW Extracellular matrix.
FT SIGNAL 1 34
FT CHAIN 35 713 ZONA PELLUCIDA SPERM-BINDING PROTEIN 2.
FT DOMAIN 35 683 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 684 703 POTENTIAL.
FT DOMAIN 704 713 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 363 630 ZP.
FT CARBOHYD 83 83 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 172 172 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 184 184 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 217 217 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 264 264 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 393 393 N-LINKED (GLCNAc. . .) (POTENTIAL).
SQ SEQUENCE 713 AA; 80209 MW; DCF9AE6CCD3461EF CRC64;

alignment_scores:
Quality: 48.50 Length: 38
Ratio: 2.553 Gaps: 2
Percent Similarity: 50.000 Percent Identity: 34.211

alignment_block:
US-09-684-016-48411 x ZP2_MOUSE ..

Align seg 1/1 to: ZP2_MOUSE from: 1 to: 713

      8 TGTAAATGAGATGCTTACAAACCTGATGACTTAATGGCGCAATGCA 57
      ||| ||||| ||| ::| ::| ||||| ||||| ::|
353 CysHsCysGluSerProValSerIleAspGluLeuGlyAlaGln...As 368
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
58 GGGCGCACTGACGTGTAAGTTACATGTAAAGTATGATCTTACTTCA 107
      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
368 pGlyPheMetAspPhe.....GluV 375
      ::||| |||
108 TTATTCACACAA 121
      ::||| |||

```

```

375 alyrSerHisGln 379
seq_name: SwissProt_39:ITB2_BOVIN
seq_documentation_block:
ID ITB2_BOVIN STANDARD; PRT; 769 AA.
AC P32592;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE INTEGRIN BETA-2 PRECURSOR (CELL SURFACE ADHESION GLYCOPROTEINS LFA-
DE 1/CD3/150.95 BETA-SUBUNIT) (CD18) (COMPLEMENT RECEPTOR C3 BETA-
DE SUBUNIT).
GN ITGB2 OR CD18.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxId=9913;
[1]
RN
RP SEQUENCE FROM N.A.
RX MEDLINE-92290287; PubMed-1351021;
RA Shuster D.E., Bosworth B.T., Kehrl M.E. Jr.:
RT "Sequence of the bovine CD18-encoding cDNA: comparison with the human
RT and murine glycoproteins."
RL Gene 114:267-271(1992).
[2]
RN
RP VARIANT LAD GLY-128.
RX MEDLINE-93028437; PubMed-1384046;
RA Shuster D.E., Kehrl M.E. Jr., Ackermann M.R., Gilbert R.O.:
RT "Identification and prevalence of a genetic defect that causes
RT leukocyte adhesion deficiency in Holstein cattle."
RL Proc. Natl. Acad. Sci. U.S.A. 89:9225-9229(1992).
-1- FUNCTION: INTEGRIN ALPHA-L/BETA-2 IS A RECEPTOR FOR ICAM1, ICAM2,
-1- ICAM3 AND ICAM4. INTEGRINS ALPHA-M/BETA-2 AND ALPHA-X/BETA-2 ARE
-1- RECEPTORS FOR THE IC3B FRAGMENT OF THE THIRD COMPLEMENT COMPONENT
-1- AND FOR FIBRINOGEN. INTEGRIN ALPHA-X/BETA-2 RECOGNIZES THE
-1- SEQUENCE G-P-R IN FIBRINOGEN ALPHA-CHAIN. INTEGRIN ALPHA-M/BETA-2
-1- RECOGNIZES P1 AND P2 PEPTIDES OF FIBRINOGEN GAMMA CHAIN. INTEGRIN
-1- ALPHA-M/BETA-2 IS ALSO A RECEPTOR FOR FACTOR X. INTEGRIN ALPHA-
-1- D/BETA-2 IS A RECEPTOR FOR ICAM3 AND VCAM1.
-1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT. BETA-2
-1- ASSOCIATES WITH EITHER ALPHA-L, ALPHA-M, ALPHA-X OR ALPHA-D.
-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
-1- DISEASE: DEFECTS IN ITGB2 ARE THE CAUSE OF LEUKOCYTE ADHESION
-1- DEFICIENCY (LAD). THE MUTATION CAUSING LAD (GLY-128) IS PREVALENT
-1- AMONG HOLSTEIN CATTLE THROUGHOUT THE WORLD, PLACING THIS DISORDER
-1- AMONG THE MOST COMMON GENETIC DISEASES KNOWN IN ANIMAL
-1- AGRICULTURE. ALL CATTLE WITH THE MUTANT ALLELE ARE RELATED TO ONE
-1- BULL, WHO THROUGH THE USE OF ARTIFICIAL INSEMINATION STRED MANY
-1- CALVES IN THE 1950S AND 1960S.
-1- SIMILARITY: BELONGS TO THE INTEGRIN BETA CHAIN FAMILY.
-1- SIMILARITY: CONTAINS 1 VWFA-LIKE DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M81233; AAA0438.1; -.
DR PIR; JC1121; JC1121.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002369; Integrin_B.
DR InterPro; IPR001169; Integrin_beta_C.
DR InterPro; IPR003659; PSI.
DR InterPro; IPR002035; VWFA.
DR Pfam; PF00362; Integrin_B_1.
DR PRINTS; PR01186; INTEGRINB.
DR PRODOM; PD001811; Integrin_B_1.
DR SMART; SM00187; INB; 1.

```


Quality: 48.00 Length: 27
Ratio: 3.429 Gaps: 1
Percent Similarity: 51.852 Percent Identity: 33.333

alignment_block:

US-09-684-016-48411 x DMSB_HAEIN ..

Align seg 1/1 to: DMSB_HAEIN from: 1 to: 205

2 AGCATGTGTAATGTGAGATGCCCTTACACCCCTGATGACCTA..... 43

103 ArgTyrCysHisMetAlaCysProTyrAspAlaProGlnTyrAspAlaG1 119

44ATGGCCATCGAGGAGTCG 64

119 nLysGlyHisMetThrTyrCysAspGlyCys 129

seq_name: SwissProt_39:VIPR_MELGA

seq_documentation_block:

ID VIPR_MELGA STANDARD; PRT; 260 AA.

AC 091085; 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DE 15-JUL-1998 (Rel. 36, Last annotation update)

DE VASOACTIVE INTESTINAL POLYPEPTIDE RECEPTOR (VIP-R) (VIP RECEPTOR)

OS Melagris gallopavo (Common turkey).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Meleagridae; Meleagris.

NCBI_TaxID=9103;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Small intestine;

RA MEDLINE=96206340; PubMed=8618952;

RA Xu M., Proudman J.A., Pits G.R., Wong E.A., Foster D.N.,

"Vasoreactive intestinal peptide stimulates prolactin mRNA expression

in turkey pituitary cells: effects of dopaminergic drugs.";

Proc. Soc. Exp. Biol. Med. 212:52-62(1996).

-1- FUNCTION: THIS IS A RECEPTOR FOR VIP. THE ACTIVITY OF THIS

RECEPTOR IS MEDIATED BY G PROTEINS WHICH ACTIVATE ADENYLYL

CYCLASE.

-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.

-1- SIMILARITY: BELONGS TO FAMILY 2 OF G-PROTEIN COUPLED RECEPTORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL: U31991; AAA99740.1; -

DR GCRB: GCR_1749; -

DR InterPro: IPR000832; GPCR_secretin.

DR Pfam: PR00002; 7tm_2; 1

DR PROSITE: PS00649; G_PROTEIN_RECP_F2_1; PARTIAL.

KW PROSITE: PS00650; G_PROTEIN_RECP_F2_2; 1.

KW G-protein coupled receptor; Transmembrane; Glycoprotein.

FT NON_TER 1 1

FT DOMAIN 1 18 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 19 42 3 (POTENTIAL).

FT TRANSMEM 43 56 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 57 78 4 (POTENTIAL).

FT TRANSMEM 79 95 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 96 119 5 (POTENTIAL).

FT DOMAIN 120 144 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 145 164 6 (POTENTIAL).

FT DOMAIN 165 176 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 177 196 7 (POTENTIAL).

FT DOMAIN 197 260 CYTOPLASMIC (POTENTIAL).

SEQ SEQUENCE 260 AA; 30358 MW; 15761E6AB5B23D5A CRC64;

alignment_scores:

Quality: 48.00 Length: 23

Ratio: 3.000 Gaps: 1

Percent Similarity: 69.565 Percent Identity: 39.130

alignment_block:

US-09-684-016-48411/rev x VIPR_MELGA ..

Align seg 1/1 to: VIPR_MELGA from: 1 to: 260

75 TACACGTCAGTGCAGCCCTGCGATTCGCGCATTAGTCATGAGGTGTA 26

2 PheGlnSerGlyGlnProGlnHisCysPheValSerSer.ValGlyCysL 18

25 AGGCATCTCACATTTC 9

18 ysaIaMetMetValpHe 23

seq_name: SwissProt_39:TFZ_HUMAN

seq_documentation_block:

ID TFZ_HUMAN STANDARD; PRT; 292 AA.

AC 016635; 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE TAFAZZIN.

GN TAZ OR EFE2 OR G4.5.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=skeletal muscle, and Heart;

RA MEDLINE=96224398; PubMed=8630491;

RA Blone S., D'Adamo P., Maestrini E., Gedeon A.K., Bolhuis P.A.,

Tonolo D.;

"A novel X-linked gene, G4.5, is responsible for Barth syndrome.";

Nat. Genet. 12:385-389(1996).

-1- SUBCELLULAR LOCATION: ISOFORMS WITH HYDROPHOBIC N-TERMINUS ARE

THOUGHT TO BE MEMBRANE ANCHORED. SHORTEST FORMS, LACKING THE

HYDROPHOBIC STRETCH, MAY BE SOLUBLE CYTOPLASMIC PROTEINS.

CC WITH THE SECOND METHIONINE AT POS. 25. ALTERNATIVE SPLICING OF THE

CC HYDROPHILIC CENTRAL REGION (EXONS 5-7) MAY PRODUCE 5 VARIANTS. TWO

CC ISOFORMS, CONTAINING ALL THREE EXONS OR LACKING EXON 5 ONLY ARE

CONSTANTLY ABUNDANT. THE SEQUENCE SHOWN HERE IS THAT OF THE

LONGEST ISOFORM.

-1- TISSUE SPECIFICITY: HIGH LEVELS IN CARDIAC AND SKELETAL MUSCLE. UP

CC TO 10 ISOFORMS CAN BE PRESENT IN DIFFERENT AMOUNTS IN DIFFERENT

CC TISSUES. MOST ISOFORMS ARE UBIQUITOUS. ISOFORMS THAT LACK THE N-

CC TERMINUS ARE FOUND IN LEUKOCYTES AND FIBROBLASTS, BUT NOT IN HEART

CC AND SKELETAL MUSCLE. SOME FORMS APPEAR RESTRICTED TO CARDIAC AND

CC SKELETAL MUSCLE OR TO LEUKOCYTES.

-1- DOMAIN: THE HYDROPHILIC DOMAIN MAY SERVE AS AN EXPOSED LOOP

CC INTERACTING WITH OTHER PROTEINS.

-1- DISEASE: DEFECTS IN TAZ ARE THE CAUSE OF BARTH SYNDROME (BTHS), A

CC SEVERE INHERITED DISORDER, OFTEN FATAL IN CHILDHOOD, CHARACTERIZED

CC BY CARDIAC AND SKELETAL MYOPATHY, SHORT STATURE AND NEUTROPENIA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL: X92763; CAA63419.1; -

DR EMBL: X92764; CAA63419.1; JOINED.
 DR EMBL: X92762; CAA63418.1; -.
 DR MIM: 302060; -.
 DR MIM: 300069; -.
 DR InterPro: IPR002123; Acyltransferase.
 DR InterPro: IPR000872; Tafazzin.
 DR Pfam: PF01553; Acyltransferase; 1.
 DR PRINTS: PR00979; TAFAZZIN.
 DR Alternative splicing: Transmembrane.
 KW TRANSMEM 1 30 MEMBRANE ANCHORED (POTENTIAL).
 FT DOMAIN 124 194 HYDROPHILIC.
 FT VARSPIC 124 153 MISSING (IN ISOFORMS WITHOUT EXON 5).
 FT VARSPIC 154 180 MISSING (IN ISOFORMS WITHOUT EXON 6).
 FT VARSPIC 181 194 MISSING (IN ISOFORMS WITHOUT EXON 7).
 SQ SEQUENCE 292 AA; 33459 MW; 2805E0962DC4FE52 CRC64;

alignment_scores:
 Quality: 48.00 Length: 55
 Ratio: 1.920 Gaps: 2
 Percent Similarity: 45.455 Percent Identity: 23.636

alignment_block:

US-09-684-016-48411/rev x TFZ_HUMAN ..

Align seg 1/1 to: TFZ_HUMAN from: 1 to: 292

```

116 TCGAATTAATTGAGATAGTATCACCCTTACAT..... 84
|||||
35 TPTPLHYSSTYMetLsnHslLeuThrValHlsAsnArgGluValLeuTy 51
|||||
83 .....GTGAACCTTACCAGTCAGTCAGCCCTCCCATTCGCCCATTAG 41
|||||
51 rGluleuLeuGluLysArgGlyProAlaThrProLeuIleThrValSerA 68
|||||
40 GTCATCAGGGTGT.....AAGCA 21
|||||
68 snHslGlnSerCysMetAspAspProHslLeuTrpGluLeuLysLeu 84
|||||
20 TCTCATTTCAT 6
|||||
85 ArgHslIleTrpAsn 89

```

seq_name: SwissProt_39:VA0D_MANSE

seq_documentation_block:

ID VA0D_MANSE STANDARD; PRT; 348 AA.
 AC Q25531;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE VACUOLAR ATP SYNTHASE SUBUNIT D (EC 3.6.1.34) (V-ATPASE D SUBUNIT)
 DE (VACUOLAR PROTON PUMP D SUBUNIT) (V-ATPASE 40 KDA SUBUNIT) (M40).
 OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
 OC Plekaryota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
 OC Sphingioidea; Sphingidae; Sphinginae; Manduca.
 NC NCBITaxid=7130;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97415414; PubMed=9271213;
 RA Merzendorf H., Harvey W.R., Wiczorek H.;
 RT "Sense and antisense RNA for the membrane associated 40 kDa subunit
 RT M40 of the insect V-ATPase.";
 RL FEBS Lett. 411:239-244(1997).
 CC -1- FUNCTION: SUBUNIT OF THE INTEGRAL MEMBRANE V0 COMPLEX OF VACUOLAR
 CC ATPASE. VACUOLAR ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY OF
 CC INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS, THUS PROVIDING
 CC MOST OF THE ENERGY REQUIRED FOR TRANSPORT PROCESSES IN THE
 CC VACUOLAR SYSTEM.
 CC -1- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A
 CC PERIPHERAL CATALYTIC V1 COMPLEX (COMPONENTS A TO H) ATTACHED TO
 CC AN INTEGRAL MEMBRANE V0 PROTON PORE COMPLEX (COMPONENTS: A, C, C',

CC C' AND D).
 CC -1- SIMILARITY: BELONGS TO THE V-ATPASE V0D/AC39 SUBUNIT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).
 CC -----
 DR EMBL: X98825; CAA67343.1; -.
 DR InterPro: IPR002843; VAMP-synt_AC39.
 DR Pfam: PF01992; VAMP-synt_AC39; 1.
 KW Hydrolase; Hydrogen ion transport.
 SQ SEQUENCE 348 AA; 39596 MW; 016BD4600B6EF098 CRC64;

alignment_scores:
 Quality: 48.00 Length: 42
 Ratio: 2.182 Gaps: 1
 Percent Similarity: 52.381 Percent Identity: 30.952

alignment_block:

US-09-684-016-48411/rev x VA0D_MANSE ..

Align seg 1/1 to: VA0D_MANSE from: 1 to: 348

```

107 TTGAAGTAGTATCCCTTACATGTGAACCTTACCAGTCAGTCAGCC 58
|||||
108 LeuLeuIleThrGluThrLeuHslGlnArgProIleSerGluLeuLepr 124
|||||
57 TCGCATTCGCCCATTAGT.....CATCAG 32
|||||
124 oLysCysHslProLeuGlySerPheGluGlnMetGluAlaIleHslValA 141
|||||
31 GTTGTAAGCATCTCATTTCAT 6
|||||
141 LaAlaThrProAlaGluLeuTyAsn 149

```

seq_name: SwissProt_39:LHX9_HUMAN

seq_documentation_block:

ID LHX9_HUMAN STANDARD; PRT; 388 AA.
 AC Q9NO69; O9NO70;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DE LIM/HOMEOBOX PROTEIN LHX9.
 DE LIM/HOMEOBOX PROTEIN LHX9.
 GN LHX9.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NC NCBITaxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Ottolenghi C.;
 RT "Absence of mutations involving the gene LHX9 in 46,XY gonadal
 RT agenesis and dysgenesis.";
 RL Submitted (May-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: INVOLVED IN GONADAL DEVELOPMENT (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -1- SIMILARITY: TO OTHER HOMEOBOX DOMAINS. BELONGS TO THE LIM
 CC SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 2 LIM DOMAINS. THE LIM DOMAIN BINDS 2 ZINC
 CC IONS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

RT and retinoic acid of the rat nerve growth factor receptor promoter.";
 RL Gene 121:247-254(1992).
 RN [3]
 RP STRUCTURE BY NMR OF 334-418
 RX MEDLINE-97449145; PubMed-9305641;
 RA Liepinsh E., Ilag L., Orling G., Ibanez C.F.;
 RT "NMR structure of the death domain of the p75 neurotrophin receptor.";
 RL EMBL J. 16:4999-5005(1997).
 CC -1- FUNCTION: LOW AFFINITY RECEPTOR WHICH CAN BIND TO NGF, BDNF,
 CC NT-3, AND NT-4.
 CC -1- SUBUNIT: NGF RECEPTOR CAN FORM A HOMODIMER THROUGH DISULFIDE
 CC BOND FORMATION.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- PTM: N- AND O-GLYCOSYLATED AND IS PHOSPHORYLATED ON SERINE.
 CC -1- SIMILARITY: CONTAINS A LA-NGFR/TNFR-TYPE CYSTEINE-RICH REGION.
 CC -1- SIMILARITY: CONTAINS 1 DEATH DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X05137; CAA28783.1; -
 CC EMBL: X61269; -; NOT_ANNOTATED_CDS.
 CC PIR: A26431; A26431.
 CC PDB: INRG; 29-70U-97.
 CC InterPro: IP000488; Death.
 CC InterPro: IP001368; TNFR_c6.
 CC Pfam: PF00531; death; 1.
 CC Pfam: PF00020; TNFR_c6; 4.
 CC SMART: SM00005; DEATH; 1.
 CC SMART: SM00208; TNFR; 3.
 CC PROSITE: PS00652; TNFR_NGFR_1; 3.
 CC PROSITE: PS50050; TNFR_NGFR_2; 4.
 CC DR PROSITE: PS50017; DEATH_DOMAIN; 1.
 CC KW Receptor; Neurogenesis; Transmembrane; Glycoprotein; Repeat;
 CC Phosphorylation; Signal; 3D-structure.
 CC FT SIGNAL 1 29
 CC FT CHAIN 30 425 LOW-AFFINITY NERVE GROWTH FACTOR
 CC FT DOMAIN 30 251 RECEPTOR.
 CC FT TRANSMEM 252 273 EXTRACELLULAR (POTENTIAL).
 CC FT DOMAIN 274 425 POTENTIAL.
 CC FT DOMAIN 32 190 CYTOPLASMIC (POTENTIAL).
 CC FT REPEAT 32 66 4 X TNFR-CYS.
 CC FT REPEAT 67 108 TNFR-CYS 1.
 CC FT REPEAT 109 148 TNFR-CYS 2.
 CC FT REPEAT 149 190 TNFR-CYS 3.
 CC FT DOMAIN 198 249 TNFR-CYS 4.
 CC FT DOMAIN 354 419 SER/THR-RICH.
 CC FT DISULFID 33 44 DEATH.
 CC FT DISULFID 45 58 BY SIMILARITY.
 CC FT DISULFID 48 65 BY SIMILARITY.
 CC FT DISULFID 68 84 BY SIMILARITY.
 CC FT DISULFID 87 100 BY SIMILARITY.
 CC FT DISULFID 90 108 BY SIMILARITY.
 CC FT DISULFID 110 123 BY SIMILARITY.
 CC FT DISULFID 126 139 BY SIMILARITY.
 CC FT DISULFID 129 147 BY SIMILARITY.
 CC FT DISULFID 150 165 BY SIMILARITY.
 CC FT DISULFID 168 181 BY SIMILARITY.
 CC FT DISULFID 171 189 BY SIMILARITY.
 CC FT CARBOHYD 61 61 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC FT CARBOHYD 71 71 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC SEQUENCE 425 AA; 45432 MW; B2E15D94D3827F8 CRC64;

alignment_scores: 48.00 Length: 20
 Quality: 3.692 Gaps: 0
 Ratio: 3.692

Percent Similarity: 65.000 Percent Identity: 35.000
 alignment_block:
 US-09-684-016-48411 x NGFR_RAT ..
 Align seg 1/1 to: NGFR_RAT from: 1 to: 425
 8 TGTAAATGTGAGATGCCCTTACACCCGTGATGACCTAATGCGCAATGCGA 57
 |||::||| ||| |||::: |||
 108 GysarGcysAlaTyrglyTyrTyrGlnAspGlnuGlnuThrglyHiscysGsl 124
 58 GGCGTGCACCT 67
 |||::|||:::
 124 uAlacysSer 127
 seq_name: SwissProt_39:GAC3_HUMAN
 seq_documentation_block:
 ID GAC3_HUMAN STANDARD; PRT; 467 AA.
 AC 099928; Q9HD46; Q9NYT2;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE GAMMA-AMINOBUTYRIC-ACID RECEPTOR GAMMA-3 SUBUNIT PRECURSOR (GABA(A)
 DE RECEPTOR).
 GN GABRG3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Fetal brain;
 RX MEDLINE-96360042; PubMed-8719414;
 RA Hadingham K.L., Walford K.A., Thompson S.A., Palmer K.J.,
 RA Whiting P.J.;
 RT "Expression and pharmacology of human GABA(A) receptors containing
 RT gamma 3 subunits.";
 RL Eur. J. Pharmacol. 291:301-309(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Han M.K., Chen Y.-H., Nurni E.L., Sutcliffe J.S.;
 RT "Gene structure and organization of the GABA(A) receptor gamma 3 subunit
 RT (GABRG3) gene.";
 RN Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 19-467 FROM N.A.
 RA Kim S.-Y., Gonen D., Yang Z.-Y., Reliford A., Dy A., Leventhal B.L.,
 RA Cook E.H., Jr.;
 RT "Genomic organization of GABA(A) receptor gamma 3 subunit gene
 RT (GABRG3)." to the EMBL/GenBank/DBJ databases.
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: GABA, THE MAJOR INHIBITORY NEUROTANSITTER IN THE
 CC VERTEBRATE BRAIN, MEDIATES NEURONAL INHIBITION BY BINDING TO THE
 CC GABA/BENZODIAZEPINE RECEPTOR AND OPENING AN INTEGRAL CHLORIDE
 CC CHANNEL.
 CC -1- SUBUNIT: GENERALLY PENTAMERIC. THERE ARE FIVE TYPES OF GABA(A)
 CC RECEPTOR CHAINS: ALPHA, BETA, GAMMA, DELTA, AND RHO.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- MISCELLANEOUS: THIS SUBUNIT CARRIES THE BENZODIAZEPINE BINDING
 CC SITE.
 CC -1- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: S82769; AAB39369.1; -
 CC EMBL: AF269144; AAF99698.1; -

OS Homo sapiens (Human)

```

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A. (LONG ISOFORM).
RC TISSUE=Monocytic Leukemia;
RX MEDLINE=99402951; PubMed=10471807;
RA Kajita M., Kinoh H., Ito N., Takamura A., Itoh Y., Okada A., Sato H.,
RA Seki M.;
RT "Human membrane type-4 matrix metalloproteinase (MT4-MMP) is encoded
RT by a novel major transcript: isolation of complementary DNA clones
RT for human and mouse mt4-mmp transcripts.";
RL FEBS Lett. 457:353-356(1999).
RN [2]
RP SEQUENCE FROM N.A. (SHORT ISOFORM).
RC TISSUE=Breast carcinoma;
RX MEDLINE=96234364; PubMed=8640782;
RA Puente X.S., Pendas A.M., Llano E., Velasco G., Lopez-Otin C.;
RT "Molecular cloning of a novel membrane-type matrix metalloproteinase
RT from a human breast carcinoma.";
RN Cancer Res. 56:944-949(1996).
RN [3]
RP SEQUENCE OF 129-302 FROM N.A. AND CHARACTERIZATION.
RX MEDLINE=20020281; PubMed=10551873;
RA Wang Y., Johnson A.R., Ye Q.-Z., Dyer R.D.;
RT "Catalytic activities and substrate specificity of the human membrane
RT type 4 matrix metalloproteinase catalytic domain.";
RN J. Biol. Chem. 274:33043-33049(1999).
RN [4]
RP GPI-ANCHOR
RX MEDLINE=20036570; PubMed=10567400;
RA Itoh Y., Kajita M., Kinoh H., Mori H., Okada A., Seki M.;
RT "Membrane type 4 matrix metalloproteinase (MT4-MMP, MMP-17) is a
RT glycosylphosphatidylinositol-anchored proteinase.";
RN J. Biol. Chem. 274:34260-34266(1999).
RN [5]
RP CHARACTERIZATION
RX MEDLINE=20006793; PubMed=10543448;
RA Kokenrock H., Essers L., Udrlich N., Will H.;
RT "Biochemical characterization of the catalytic domain of membrane-type
RT 4 matrix metalloproteinase.";
RN J. Biol. Chem. 274:1103-1108(1999).
CC -1- FUNCTION: ENOPEPTIDASE THAT DEGRADES VARIOUS COMPONENTS OF THE
CC EXTRACELLULAR MATRIX, SUCH AS FIBRIN. MAY BE INVOLVED IN THE
CC ACTIVATION OF MEMBRANE-BOUND PRECURSORS OF GROWTH FACTORS OR
CC INFLAMMATORY MEDIATORS, SUCH AS TUMOR NECROSIS FACTOR ALPHA. MAY
CC ALSO BE INVOLVED IN TUMORAL PROCESS. NOT OBVIOUS IF ABLE TO
CC PROTEOLYTICALLY ACTIVATES PROGELATINASE A. DOES NOT HYDROLYSE
CC COLLAGEN TYPES I, II, III, IV AND V, GELATIN, FIBRONECTIN,
CC LAMININ, DECORIN NOR ALPHA1-ANTITRYPSIN.
CC -1- CATALYTIC ACTIVITY: CLEAVES PRO-TNF-ALPHA AT THE 74-ALA-|-GLN-75
CC SITE.
CC -1- COFACTOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT/PUENTE-TYPE FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LEUKOCYTES, COLON, OVARY
CC TESTIS AND BREAST CANCER. EXPRESSED ALSO IN MANY TRANSFORMED AND
CC NON-TRANSFORMED CELL TYPES.
CC -1- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 HEMOPEXIN-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A (ZINC
CC METALLOPROTEASE) ALSO KNOWN AS MATRIXIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```


RT	seq_documentation_block:	PRT;	325 AA.
ID	EGIP_ANTCR STANDARD;		
AC	P15217; P15218; P15219;		
DC	01-APR-1990 (Rel. 14, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	01-OCT-1996 (Rel. 34, Last annotation update)		
DE	EXOCASTRULA-INDUCING POLYPEPTIDE PRECURSOR (EGIP).		
OS	Anthicodiaris crassispina (Sea urchin).		
OC	Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;		
OC	Echinioidea; Euechinioidea; Echinacea; Echinoida; Echinometridae;		
OX	NCBI_TaxId=7629;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Embryo.		
TX	MEDLINE=95220383; PubMed=7705369;		
RA	Yamasu K., Watanabe H., Kohchi C., Soma G.-I., Mizuno D.-I.,		
RA	Akasaka K., Shimada H., Suyemitsu T., Ishihara K.;		
RT	"Molecular cloning of a cDNA that encodes the precursor to several		
RT	extracellular inducing peptides, epidermal growth factor-related		

RT polypeptides of the sea urchin *Anthocidaris crassispina*.
 RL Eur. J. Biochem. 228:515-523(1995).
 RN [2]
 RP SEQUENCE OF 106-158 AND 179-230.
 RC TISSUE=Embryo;
 RX MEDLINE=89230051; PubMed=2713739;
 RA Suyemitsu T., Asami-Yoshizumi T., Noguchi S., Tonegawa Y.,
 RT Ishihara K.;
 RT "The exogastrula-inducing peptides in embryos of the sea urchin,
 RT *Anthocidaris crassispina* -- isolation and determination of the primary
 RT structure."
 RL Cell Differ. Dev. 26:53-66(1989).
 RN [3]
 RP SEQUENCE OF 47-104.
 RC TISSUE=Embryo;
 RX MEDLINE=90028385; PubMed=2804137;
 RA Suyemitsu T., Tonegawa Y., Ishihara K.;
 RT "Amino acid sequence of exogastrula-inducing peptide C from the sea
 RT urchin, *Anthocidaris crassispina*."
 RL Biochim. Biophys. Acta 993:24-28(1989).
 CC -1- FUNCTION: THE EGIP PEPTIDES ARE FACTORS EFFECTIVE TO EXTRUDE THE
 CC ARCHENTERON TOWARD OUTSIDE OF EMBRYOS. MAY HAVE A ROLE IN THE
 CC INDUCTION OF GASTRULATION.
 CC -1- SIMILARITY: EACH OF THE 4 PEPTIDES CONTAINS 1 EGF-LIKE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z48184; CA88234.1; -
 DR PIR; A48167; A48167.
 DR PIR; B48167; B48167.
 DR PIR; S06637; S06637.
 DR HSSP; P00592; 3P2P.
 DR InterPro; IPR000561; EGF-like.
 DR Pfam; PF00008; EGF; 2.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00001; EGF-like; 2.
 DR PROSITE; PS00186; EGF_1; 4.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 KW Signal; Developmental protein; EGF-like domain; Extracellular matrix;
 KW Repeat; Gastrulation.
 FT SIGNAL 1 19 POTENTIAL.
 FT PEPTIDE 47 104 EXOGASTRULA-INDUCING PEPTIDE C (EGIP-C).
 FT PEPTIDE 106 158 EXOGASTRULA-INDUCING PEPTIDE D (EGIP-D).
 FT PEPTIDE 179 230 EXOGASTRULA-INDUCING PEPTIDE A (EGIP-A).
 FT PEPTIDE 251 310 EGIP-X (PROBABLE).
 FT DOMAIN 48 91 EGF-LIKE 1.
 FT DOMAIN 107 154 EGF-LIKE 2.
 FT DOMAIN 180 226 EGF-LIKE 3.
 FT DOMAIN 252 298 EGF-LIKE 4.
 FT DISULFID 52 65 BY SIMILARITY.
 FT DISULFID 59 75 BY SIMILARITY.
 FT DISULFID 77 90 BY SIMILARITY.
 FT DISULFID 111 124 BY SIMILARITY.
 FT DISULFID 118 138 BY SIMILARITY.
 FT DISULFID 140 153 BY SIMILARITY.
 FT DISULFID 184 197 BY SIMILARITY.
 FT DISULFID 191 211 BY SIMILARITY.
 FT DISULFID 213 225 BY SIMILARITY.
 FT DISULFID 256 269 BY SIMILARITY.
 FT DISULFID 263 283 BY SIMILARITY.
 FT DISULFID 285 297 BY SIMILARITY.
 SQ SEQUENCE 325 AA; 36462 MW; 1FD2577B3617306A CRC64;

alignment_scores:

Quality: 47.50
 Ratio: 3.393

Length: 39
 Gaps: 1

Percent Similarity: 35.897 Percent Identity: 25.641
 alignment_block:
 US-09-684-016-48411 x EGIP_ANTCR ..
 Align seg 1/1 to: EGIP_ANTCR from: 1 to: 325
 5 TATGTAAATGTAGATGCTTAC..... 28
 |||||.....
 74 TycCysCysCysThrLeuProTYrArgValGlySerCluserSercy 90
 29AACCCGATGACC 41
 90 sTYrMetProLysaspLysGluaspValGluIleGluThrLysaspr 107
 42 TAATGGCGCAATGCGAG 58
 :|||:.....
 107 hrValAlaArgCysGlu 112

GE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

OM of: US-09-684-016-48411 to: SPTREMBL_17:* out_format: pfs

Date: Mar 19, 2002 9:34 AM

About: Results were produced by the GenCore software, version 4.5,

Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODE=frame:r2p.model -DEV=xlh
-QC/cgn2.1/USPTO_SPOOL/US09684016/runat.19032002.084440.20931/app.query.fasta.1.182
-DB=SPTREMBL_17 -QFMT=fastan -SUFFIX=std.rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000
-OCAPOP=4.500 -OCAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCAPOP=6.000 -FCAPEXT=7.000 -YCAPOP=10.000 -YCAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human0.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NORMEXT -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09684016.cgn1.1.96 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPPY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-684-016-48411
Query Length: 123
Database: SPTREMBL_17:*
Database sequences: 473505
Database length: 146272329
Search time (sec): 65.740000
```

Score list:

Sequence	Strd	Orig	ZScore	E_Score	Len	Documentation
SP_plant:065462	+	138.00	374.72	6.7e-13	201	065462 arabidopsis thaliana (mc
SP_plant:043479	+	127.00	342.84	3.3e-11	227	043479 hordeum vulgare (barley)
SP_plant:09F9E9	+	124.00	334.34	1.0e-10	228	09F9E9 arabidopsis thaliana (mc
SP_plant:09SV14	+	124.00	333.51	1.0e-10	258	09SV14 arabidopsis thaliana (mc
SP_plant:09XJ51	+	98.00	268.78	1.5e-06	71	09XJ51 triticum aestivum (wheat)
SP_rdent:09WTU0	+	77.00	191.06	0.0021	1096	09WTU0 mus musculus (mouse)
SP_human:09Y6M4	+	77.00	191.04	0.0021	1099	09Y6M4 homo sapiens (human)
SP_invertebrate:09VBE3	+	76.50	196.37	0.0029	404	09VBE3 drosophila melanogaster
SP_human:09H4H2	+	74.00	184.97	0.0065	790	09H4H2 homo sapiens (human)
SP_human:09UPE1	+	74.00	182.66	0.0062	1084	09UPE1 homo sapiens (human)
SP_fungi:003012	+	66.50	169.04	0.1089	353	003012 saccharomyces cerevisiae
SP_fungi:074508	+	62.50	156.51	0.4521	424	074508 schizosaccharomyces pombe
SP_invertebrate:045410	+	62.00	155.41	0.5453	405	045410 caenorhabditis elegans
SP_invertebrate:09N1C1	+	62.00	153.86	0.5284	451	09N1C1 caenorhabditis elegans
SP_invertebrate:045407	+	60.00	136.79	0.8652	2781	045407 caenorhabditis elegans
SP_human:09U1G2	+	60.00	132.90	0.7996	4957	09U1G2 homo sapiens (human)
SP_human:014687	+	60.00	132.49	0.7931	5262	014687 homo sapiens (human)
SP_invertebrate:022662	+	59.50	148.17	1.134	416	022662 caenorhabditis elegans
SP_plant:09T0A1	+	59.50	147.75	1.133	443	09T0A1 arabidopsis thaliana (mc
SP_invertebrate:09VRV6	+	57.50	139.39	2.260	662	09VRV6 drosophila melanogaster
SP_fungi:09U719	+	57.00	132.09	2.77	1588	09U719 schizosaccharomyces pom
SP_human:09BQW0	+	56.50	140.33	4.03	379	09BQW0 homo sapiens (human)
SP_human:09BNC3	+	56.50	137.89	3.84	544	09BNC3 homo sapiens (human)
SP_human:09N1M8	+	56.50	137.67	3.82	562	09N1M8 homo sapiens (human)
SP_human:09BNC0	+	56.50	137.67	3.82	562	09BNC0 homo sapiens (human)
SP_rdent:09WV00	+	56.50	137.08	3.78	614	09WV00 mus musculus (mouse)
SP_invertebrate:09W664	+	56.50	136.79	3.75	991	09W664 barbus (barbel)
SP_human:015043	+	56.50	133.85	3.54	991	015043 homo sapiens (human)
SP_human:09H4G6	+	56.50	133.52	3.51	1042	09H4G6 homo sapiens (human)
SP_archaea:09YEG2	+	56.00	136.48	4.60	544	09YEG2 aeropyrum pernix (hypoc
SP_bacteria:09RLE4	+	55.50	137.79	5.83	363	09RLE4 legionella pneumophila
SP_human:09NR48	+	55.50	123.64	4.37	2969	09NR48 homo sapiens (human)
SP_invertebrate:09XMD7	+	55.00	141.11	7.68	180	09XMD7 homo sapiens (human)
SP_invertebrate:09XMD7	+	55.00	136.29	6.97	368	09XMD7 caenorhabditis elegans
SP_plant:09FNT5	+	55.00	126.76	5.75	1516	09FNT5 arabidopsis thaliana (m
SP_invertebrate:09U263	+	54.50	124.73	6.81	1645	09U263 caenorhabditis elegans
SP_invertebrate:044498	+	54.50	124.75	6.80	1655	044498 caenorhabditis elegans
SP_invertebrate:09VR08	+	54.50	119.37	6.10	3680	09VR08 drosophila melanogaster
SP_invertebrate:09VTS1	+	54.00	129.90	9.31	625	09VTS1 drosophila melanogaster

```
SP_invertebrate:024459 + 54.00 127.77 8.92 857 024459 drosophila melanog
SP_invertebrate:09V8C2 + 54.00 126.46 8.69 1042 09V8C2 drosophila melano
SP_plant:09FNT5 + 53.50 129.11 11.30 570 09FNT5 arabidopsis thaliana
SP_invertebrate:09VGT8 + 53.50 120.60 9.52 2016 09VGT8 drosophila melano
SP_virus:098VX5 + 53.00 134.82 15.65 198 098VX5 human immunodeficien
```

seq_name: sp_plant:065462

```
seq_documentation_block:
ID 065462 PRELIMINARY; PRT; 201 AA.
AC 065462;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE RECEPTOR LIKE PROTEIN (FRAGMENT).
GN FIN20.240 OR AT4G22140.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids; II: Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN 1;
RP SEQUENCE FROM N.A.
RA Beyer M., Wedler H., Wambutt R., Bancroft I., Mewes H.W., Mayer K.,
RA Schueller C.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN 12;
RP SEQUENCE FROM N.A.
RA Wedler H., Wambutt R., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN 13;
RP SEQUENCE OF 1-0 FROM N.A.
RA Robben J., Gymnopoulos B., Volckaert G., Mewes H.W., Lemcke K.,
RA Meyer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN 14;
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022140; CAA18117.1; -
DR EMBL; AL161556; CAB79169.1; -
DR Mendel; 29749; Arabid; 2000; 29749.
DR InterPro; IPR001025; BAH.
DR InterPro; IPR000561; BGF-like.
DR InterPro; IPR001965; PHD.
DR Pfam; PF01426; BAH; 1.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00439; BAH; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS01186; BGF_2; UNKNOWN_1.
SO SEQUENCE 201 AA; 22579 MW; 4C48904BB9E61498 CMC64;
```

alignment_scores:

```
Quality: 138.00 Length: 23
Ratio: 6.273 Gaps: 0
Percent Similarity: 95.652 Percent Identity: 91.304
```

alignment_block:

```
US-09-684-016-48411 x 065462
Align seg 1/1 to: 065462 from: 1 to: 201
```

```
5 TATTGTAATGTCAGATGCTTACCAACCTGATGACCTAATGGCGCATG 54
|||||
115 TYRCYLSVCSGLMEPTOTYRASNPROASPSPLEUETVALGLNCY 131
|||||
55 CGAGGCGTCACTGACTG 73
|||||
131 SGLUCLYCSLYSASPTRP 137
seq_name: sp_plant:043479
```

```

seq_documentation_block:
ID Q43479 PRELIMINARY; PRT; 227 AA.
AC Q43479;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE (DMS576).
GN ES43.
OS Hordeum vulgare (Barley).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
OC Triticaceae; Hordeum.
OX NCBI_TaxID=4513;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, DBS576; TISSUE=LEAF;
RA Speulman E., Salamini F.;
RL Plant Sci. 106:91-98(1995).
DR EMBL; X77575; CA54682.1;
DR Mendel; 12800; Horvu; 2000; 12800.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR001025; BAH.
DR InterPro; IPR001965; PHD.
DR Pfam; PF00628; PHD; 1.
DR Pfam; PF01426; BAH; 1.
DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
DR SMART; SM00439; BAH; 1.
DR SMART; SM00249; PHD; 1.
SQ SEQUENCE 227 AA; 25741 MW; EC1A86DB0A472E24 CRC64;

alignment_scores:
Quality: 127.00 Length: 23
Ratio: 6.048 Gaps: 0
Percent Similarity: 91.304 Percent Identity: 82.609

alignment_block:
US-09-684-016-48411 x Q43479 ..
Align seg 1/1 to: Q43479 from: 1 to: 227

5 TATTGTAATGTGAGATGCCTTACACCTGATGACCTATATGGCGCAATG 54
:::|||||
147 PhcysLysCysGluMetProTyrAsnProAspAspleuMetValGlnCyl 163
||| |||
55 CGAGGCTGCACCTGACTGG 73
||| |||
163 sgluGluCysSerAspTrrp 169

seq_name: sp_plant:Q9FEN9

seq_documentation_block:
ID Q9FEN9 PRELIMINARY; PRT; 228 AA.
AC Q9FEN9;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE PUTATIVE PHD FINGER TRANSCRIPTION FACTOR.
GN SHL1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Oh M., Torisky R.S., Braam J., Altmann T., Clouse S.D.;
RT "PHD Finger Dependent Binding of SHL1 to a Specific Promoter Region of
Arabidopsis TCH4 in vitro.";
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF277453; AAG21353.1;
DR InterPro; IPR001025; BAH; 1.
DR InterPro; IPR000164; Histone_H3.

```

```

DR InterPro; IPR001965; PHD.
DR Pfam; PF01426; BAH; 1.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00439; BAH; 1.
DR SMART; SM00249; PHD; 1.
DR PROSITE; PS00322; HISTONE_H3_1; UNKNOWN_1.
SQ SEQUENCE 228 AA; 26122 MW; A800BC81E5B8390A CRC64;

alignment_scores:
Quality: 124.00 Length: 23
Ratio: 5.636 Gaps: 0
Percent Similarity: 95.652 Percent Identity: 78.261

alignment_block:
US-09-684-016-48411 x Q9FEN9 ..
Align seg 1/1 to: Q9FEN9 from: 1 to: 228

5 TATTGTAATGTGAGATGCCTTACACCTGATGACCTATATGGCGCAATG 54
:::|||||
141 PhcysLysCysGluMetProTyrAsnProAspAspleuMetValGlnCyl 157
||| |||
55 CGAGGCTGCACCTGACTGG 73
||| |||
157 sgluGluCysSerAspTrrp 163

seq_name: sp_plant:Q9SVI4

seq_documentation_block:
ID Q9SVI4 PRELIMINARY; PRT; 258 AA.
AC Q9SVI4;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE ES43 LIKE PROTEIN (FRAGMENT).
GN F19H22.200 OR ATG639100.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
RA Mayer K.F.X., Scheller C.;
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Zimmermann W., Grueneisen A., Wambutt R., Kalicki J., Wohlmann P.,
RA Smith A., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AL035679; CAB38830.1;
DR EMBL; AL161594; CAB80573.1;
DR InterPro; IPR001025; BAH.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000164; Histone_H3.
DR InterPro; IPR001965; PHD.
DR Pfam; PF01426; BAH; 1.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00439; BAH; 1.

```

DR SMART; SM00249; PHD; 1.
 DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
 DR PROSITE; PS00322; HISTONE_H3_1; UNKNOWN_1.
 FT NON_TER 258
 SO SEQUENCE 258 AA; 29675 MW; 17A2P2972B1C1E77 CRC64;

alignment_scores:
 Quality: 124.00 Length: 23
 Ratio: 5.636 Gaps: 0
 Percent Similarity: 95.652 Percent Identity: 78.261

alignment_block:
 US-09-684-016-48411 x Q9SVI4 ..

Align seg 1/1 to: Q9SVI4 from: 1 to: 258

5 TATTGTAATGTGAGATGCTTACACCGTATGACCTAATGGCGCAATG 54
 ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 141 PhecylsCysGluMetProTyrAsnProAspAspMetValGlnCy 157
 55 CGAGGCGTGCACCTGACTGG 73
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 157 sgluGlnCysSerGluTrp 163

seq_name: sp_plant:Q9XJ51

seq_documentation_block:
 ID Q9XJ51 PRELIMINARY; PRT; 71 AA.

DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ZINC-FINGER MOTIF (FRAGMENT).
 GN WSR4.
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae;
 OC Triticaceae; Triticum.
 OX NCBI_TaxID=4565;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CHINESE SPRING; TISSUE=ROOT;
 RA Nemoto Y., Kawakami N., Sasakuma T.;
 RT "Isolation of novel early salt-responding genes from wheat (Triticum
 RT aestivum L.) by differential display."
 RL Theor. Appl. Genet. 98:673-678(1999).
 DR EMBL; AB011445; BAA82157.1; -.
 DR InterPro; IPR001965; PHD.
 DR Pfam; PF00628; PHD; 1.
 DR SMART; SM00249; PHD; 1.
 KW Zinc.
 FT NON_TER 1
 SQ SEQUENCE 71 AA; 8056 MW; 9E06685BC184E423 CRC64;

alignment_scores:
 Quality: 98.00 Length: 17
 Ratio: 6.125 Gaps: 0
 Percent Similarity: 94.118 Percent Identity: 88.235

alignment_block:
 US-09-684-016-48411 x Q9XJ51 ..

Align seg 1/1 to: Q9XJ51 from: 1 to: 71

23 CCTTAAACCTGATGACCTAATGCGCAATGCGGCTGCACTGACTG 72
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1 ProTyrAsnProAspAspMetValGlnCysGlnCysLysAspTr 17
 73 G 73
 1
 17 P 17

seq_name: sp_rodent:Q9WTU0

seq_documentation_block:
 ID Q9WTU0 PRELIMINARY; PRT; 1096 AA.

DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PHD-FINGER PROTEIN.
 GN PHF2 OR GRC5.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99160472; PubMed=10051327;
 RA Hasenpusch-Theil K., Chadwick B.P., Theil T., Heath S.K.,
 RA Wilkison D.G., Frischauf A.M.;
 RT "PHF2, a novel PHD finger gene located on human chromosome 9q22."
 RL Mamm. Genome 10:294-298(1999).
 DR EMBL; AF043726; AAD21792.1; -.
 DR MGD; MGI:1338034; Phf2.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR001965; PHD.
 DR Pfam; PF00628; PHD; 1.
 DR SMART; SM00249; PHD; 1.
 DR PROSITE; PS01186; EGF_2; UNKNOWN_1.
 SQ SEQUENCE 1096 AA; 120798 MW; D12C8DC10938AF CRC64;

alignment_scores:
 Quality: 77.00 Length: 23
 Ratio: 4.529 Gaps: 0
 Percent Similarity: 73.913 Percent Identity: 43.478

alignment_block:
 US-09-684-016-48411 x Q9WTU0 ..

Align seg 1/1 to: Q9WTU0 from: 1 to: 1096

5 TATTGTAATGTGAGATGCTTACACCGTATGACCTAATGGCGCAATG 54
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 7 TyrcysValCysArgLeuProTyrAspValThrArgPheMetIleGlnCy 23
 55 CGAGGCGTGCACCTGACTGG 73
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 23 sAspAlaCysLysAspTrp 29

seq_name: sp_human:Q9Y6N4

seq_documentation_block:
 ID Q9Y6N4 PRELIMINARY; PRT; 1099 AA.

AC Q9Y6N4;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PHD-FINGER PROTEIN.
 GN GRC5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99160472; PubMed=10051327;
 RA Hasenpusch-Theil K., Chadwick B.P., Theil T., Heath S.K.,
 RA Wilkison D.G., Frischauf A.M.;
 RT "PHF2, a novel PHD finger gene located on human chromosome 9q22."
 RL Mamm. Genome 10:294-298(1999).
 DR EMBL; AF043725; AAD21791.1; -.
 DR InterPro; IPR000561; EGF-like.

```

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Striffler R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weisslock G.M., Weissbach J.,
RA Williams S.M., Wodade T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL: AE003655; AAF53607.1; -.
DR FlyBase: FBgn0032635; CG51514.1.
DR InterPro: IPR001965; PHD.
DR InterPro: IPR003126; ZnF_UBR1.
DR Pfam: PF02207; zf-UBR1_1.
DR SMART: SM00249; PHD; 1.
DR SMART: SM00396; ZnF_UBR1; 1.
DR SEQUENCE 404 AA; 45969 MW; 467FB27FE9F9E995 CRC64;

alignment_scores:
    Quality: 76.50      Length: 28
    Ratio: 4.250        Gaps: 2
Percent Similarity: 64.286   Percent Identity: 50.000

alignment_block:
US-09-684-016-48411 x Q9VJE3 ..

Align seg 1/1 to: Q9VJE3. from: 1 to: 404

5 TATGTGAATGAGATGCTTAC...AACCGTAT.....GA 39
|||||..... ||||| ...|||||
145 TyrcGlsCysGlsArgProTyrProAspProAspArgTrpValGluGI 161
40 CCTAATGGCGAATGCGAGGCTGCACGTACTG 73
:::||||| ||| |||||
161 valmetleugincysalailecysgluasptyr 172

seq_name: sp_human:Q9HAH2

seq_documentation_block:
ID Q9HAH2 PRELIMINARY; PRT; 790 AA.
AC Q9HAH2;
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 17, Last annotation update)
DE CNNA FLJ11634 FIS, CLONE HEMBA1004275, MODERATELY SIMILAR TO HOMO
DE SAPIENS PHD-FINGER PROTEIN (GRCS) MRNA (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=WHOLE EMBRYO, MAINLY HEAD;
RA Iisogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosobiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Choshi Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Matanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,
RA Yamamoto J., Nakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
RA Nimomiya K., Iwayanagi T.;
RA "NED0 human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK021696; BAB13877.1; -.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001965; PHD.
DR Pfam: PF00628; PHD; 1.
DR SMART: SM00249; PHD; 1.
DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
FT NON_TER 790
SQ SEQUENCE 790 AA; 88642 MW; D4D21D2A766C671A CRC64;

```


alignment_scores:
Quality: 74.00 Length: 23
Ratio: 4.625 Gaps: 0
Percent Similarity: 69.565 Percent Identity: 43.478

alignment_block:
US-09-684-016-48411 x Q9HAH2 ..

Align seg 1/1 to: Q9HAH2 from: 1 to: 790

```
5 TATGTGAATGTGATGACCTTACACCGATGATACCTAATGGCGCAATG 54
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7 Tyricysleucysargleuprototyryaspvalthrargphenetileglucy 23
CGAGGCTGCACTGACTGG 73
|::: ||| |||||
23 saspmctcysglinspttrp 29
```

seq_name: sp_human:Q9UPP1

seq_documentation_block:
ID Q9UPP1 PRELIMINARY; PRT: 1084 AA.

```
AC Q9UPP1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE KIAA1111 PROTEIN (FRAGMENT).
GN KIAA1111.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=93937452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K., Hirotsawa M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:197-205(1999).
DR EMBL; AB029034; BAA83063.1; -.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001965; PHD.
DR Pfam: PF00628; PHD; 1.
DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
DR SMART; SM00249; PHD; 1.
FT NON_TER 1
FT SEQUENCE 1084 AA; 120376 MW; D609EBA16BE71A3E CRC64;
```

alignment_scores:
Quality: 74.00 Length: 23
Ratio: 4.625 Gaps: 0
Percent Similarity: 69.565 Percent Identity: 43.478

alignment_block:
US-09-684-016-48411 x Q9UPP1 ..

Align seg 1/1 to: Q9UPP1 from: 1 to: 1084

```
5 TATGTGAATGTGATGACCTTACACCGATGATACCTAATGGCGCAATG 54
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 Tyricysleucysargleuprototyryaspvalthrargphenetileglucy 83
CGAGGCTGCACTGACTGG 73
|::: ||| |||||
83 saspmctcysglinspttrp 89
```

seq_name: sp_fungi:Q03012

seq_documentation_block:

```
ID Q03012 PRELIMINARY; PRT: 353 AA.
AC Q03012;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE LP18P.
GN YPL138C OR LP18C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90360986; PubMed=2167832;
RA Tzagoloff A., Capitanio N., Nobrega M.P., Gatti D.;
RT "Cytochrome oxidase assembly in yeast requires the product of COX11, a
RT homolog of the P. denitrificans protein encoded by ORF3.";
RL EMBO J. 9:2759-2764(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93140769; PubMed=8423796;
RA Brigati C., Kurtz S., Balderes D., Vidali G., Shore D.;
RT "An essential yeast gene encoding a TTAGGG repeat-binding protein.";
RL Mol. Cell. Biol. 13:1306-1314(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=93233668; PubMed=8386320;
RA Irie K., Takase M., Lee K.S., Levin D.E., Araki H., Matsumoto K.,
RA Oshima Y.;
RT "MKR1 and MKR2, which encode Saccharomyces cerevisiae mitogen-
RT activated protein kinase-kinase homologs, function in the pathway
RT mediated by protein kinase C.";
RL Mol. Cell. Biol. 13:3076-3083(1993).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=94287711; PubMed=8017104;
RA Jiang B., Brown J.L., Sheraton J., Fortin N., Bussey H.;
RT "A new family of yeast genes implicated in ergosterol synthesis is
RT related to the human oxysterol binding protein.";
RL Yeast 10:341-353(1994).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=95320205; PubMed=7597071;
RA Shani N., Watkins P.A., Valle D.;
RT "PXAL, a possible Saccharomyces cerevisiae ortholog of the human
RT adrenoleukodystrophy gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:6012-6016(1995).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=97313271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Anseorge W.,
RA Araujo R., Aparicio A., Barrell B., Bedcock K., Benes V., Bolstein D.,
RA Bowman S., Bruckner M., Carpenter J., Cherry J.M., Chung E.,
RA Churcher C., Coster F., Davis K., Davis R.W., Dietrich F.S.,
RA Delius H., DiPaolo T., Dubois E., Dusterhoft A., Duncan M., Floeth M.,
RA Fortin N., Friesen J.D., Fritz C., Goffeau A., Hall J., Heblung U.,
RA Heumann K., Hilbert H., Hillier L., Hunnicke-Smith S., Hyman R.,
RA Johnston M., Kalman S., Kleine K., Komp C., Kurdi O., Lashkari D.,
RA Lew H., Lin A., Lin D., Louis E.J., Marathe R., Messenguy F.,
RA Mewes H.W., Mitrilipati S., Moestl D., Muller-Auer S., Namath A.,
RA Newnich U., Oefner P., Pearson D., Petel F.X., Pohl T.M.,
RA Purnelle D., Schafer M., Scharfe M., Scherens B., Schramm S.,
RA Schroeder M., Sdlou A.M., Tettelin H., Urestranzu L.A., Ushinsky S.,
RA Vierendeels F., Vissers S., Voss H., Walsh S.V., Wandut R., Wang Y.,
RA Wedler E., Wedler H., Winnett E., Zhong W.W., Zollner A., Vo D.H.,
RA Hani J.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL Nature 387:0-0(0).
RN [7]
RP SEQUENCE FROM N.A.
RA Swartzman E., Viswanathan M.N., Emerick A.E., Thorner J.;
RT Submitted (Apr-1996) to the EMBL/Genbank/DBD databases.
RN [8]
```

RP SEQUENCE FROM N.A.
 RA Hall J., Depaulo T., Ahmed A., Bussey H., Fortin N., Friesen J.D.,
 RA Storms R.K., Vo D.H., Wang Y., Minnett E.;
 RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
 RN [9]
 RP SEQUENCE FROM N.A.
 RA Bussey H.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 RN [10]
 RP SEQUENCE FROM N.A.
 RA Jia Y., Cherry J.M.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U43703; AAB68222.1; -.
 DR SGD: S0006059; IPL138C.
 DR InterPro: IPR001965; PHD.
 DR Pfam: PF00628; PHD; 1.
 DR SMART: SM00249; PHD; 1.
 SQ SEQUENCE 353 AA; 41468 MW; F4B8DFE0CB358CA4 CRC64;

alignment_scores:
 Quality: 66.50 Length: 23
 Ratio: 4.156 Gaps: 1
 Percent Similarity: 69.565 Percent Identity: 47.826

alignment_block:
 US-09-684-016-48411 x Q03012 ..

Align seg 1/1 to: Q03012 from: 1 to: 353

5 TATTGTAATGAGATGCTTACACCCGATGACCTATGCGCATG 54
 ||||| |||: ||| :||| |||: ||
 24 TYRCYSILCYSLYSARPRO...ASPTRYGLYLULEMELVALGLYC 39
 55 CGAGGCGTGCCTGACTGG 73
 |::||| |||||
 39 SASPLGYCYSASPSPTR 45

seq_name: sp_fungi:074508

seq_documentation_block:
 ID 074508 PRELIMINARY; PRT; 424 AA.
 AC 074508:

DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PUTATIVE TRANSCRIPTIONAL REGULATORY PROTEIN, PHD FINGER CONTAINING.
 GN SPC594.05C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_Taxid=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972H-;
 RA Rieger M., Lyne M., Rajandream M.A., Barrell B.G.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL031523; CAA20664.1; -.
 DR InterPro: IPR001965; PHD.
 DR Pfam: PF00628; PHD; 1.
 DR SMART: SM00249; PHD; 1.
 SQ SEQUENCE 424 AA; 48683 MW; 2963605C3DFCC0B9 CRC64;

alignment_scores:
 Quality: 62.50 Length: 25
 Ratio: 4.167 Gaps: 2
 Percent Similarity: 60.000 Percent Identity: 48.000

alignment_block:
 US-09-684-016-48411 x 074508 ..

Align seg 1/1 to: 074508 from: 1 to: 424
 5 TATTGTAATGAGATGCTTACACCCGATGAC.....CTAATGCC 48
 ||||| |||: ||| :||| |||: ||
 120 TYRCYSILCYSLIN.....LSPROASPSPGLYSETRIPNETTE 133
 49 GCATGCGAGGCTGCCTGACTGG 73
 |||: ||| |||||
 133 UGLYCYSASPLGYCYSGLINSPTR 141

seq_name: sp_invertebrate:045410

seq_documentation_block:
 ID 045410 PRELIMINARY; PRT; 405 AA.
 AC 045410:
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE F26H11.3A PROTEIN.
 GN F26H11.3A.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_Taxid=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Barlow K.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]

RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans".
 RL Nature 368:32-38(1994).
 DR EMBL: Z81515; CAB04198.1; -.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001487; Bromodomain.
 DR InterPro: IPR001965; PHD.
 DR Pfam: PF00439; Bromodomain; 1.
 DR Pfam: PF00628; PHD; 2.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR PROSITE: PS50014; BROMODOMAIN_2; 1.
 DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
 DR SMART: SM00297; BROMO; 1.
 DR SMART: SM00249; PHD; 2.
 SQ SEQUENCE 405 AA; 46199 MW; DE6BE7384A5AE82F CRC64;

alignment_scores:
 Quality: 62.00 Length: 23
 Ratio: 4.429 Gaps: 0
 Percent Similarity: 60.870 Percent Identity: 34.783

alignment_block:
 US-09-684-016-48411 x 045410 ..

Align seg 1/1 to: 045410 from: 1 to: 405

5 TATTGTAATGAGATGCTTACACCCGATGACCTATGCGCATG 54
 ||||| |||: ||| :||| |||: ||
 172 TYRCYSVALCYSLINLSPSPTRIRASPHRNLSPHETRYVALGLYC 188

55 CGAGGCGTGCCTGACTGG 73

OC Rhaditidae; Peloderinae; Caenorhadtis.

RT transcription factor.";

RT transcription factor.";

DR EMBL: AB032251; BAA89208.1; -
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR01487; Bromodomain.
 DR InterPro: IPR001965; PHD.
 DR Pfam: PF00439; bromodomain; 1.
 DR Pfam: PF00628; PHD; 2.
 DR PRINTS: PR00503; BROMODOMAIN.
 DR PROSITE: PS00633; BROMODOMAIN_1; 1.
 DR PROSITE: PS0014; BROMODOMAIN_2; 1.
 DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
 DR SMART: SM00297; BROMO; 1.
 DR SMART: SM00249; PHD; 2.
 SQ SEQUENCE 2781 AA; 311210 MW; 7B3BE0C1C4C3CFBE CRC64;

alignment_scores:
 Quality: 60.00 Length: 23
 Ratio: 4.286 Gaps: 0
 Percent Similarity: 60.870 Percent Identity: 34.783

alignment_block:
 US-09-684-016-48411 x Q9UIG2 ..

Align seg 1/1 to: Q9UIG2 from: 1 to: 2781

5 TATGTAAATGAGATGCTTACAACTGATGACCTAATGGCGCAATG 54
 ||||| |||:: |||||::: |||::: |||
 2604 TYRCYSILCYSILSYSTRPOTYRASPGLUSERTYSRHYRILGLICY 2620
 55 CGAGGCGCTGCACCTGACTGG 73
 1::: |||::: |||
 2620 SASPARCYSGINASNTRP 2626

seq_name: sp_human:014687

seq_documentation_block:
 ID 014687; PRELIMINARY; PRT; 4957 AA.
 AC 014687;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ALR.
 GN ALR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 ON NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97388474; PubMed=9247308;
 RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
 RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
 RA Cnaan E.;
 RT "Structure and expression pattern of human ALR, a novel gene with
 RT strong homology to ALL-1 involved in acute leukemia and to Drosophila
 RT trithorax.";
 RL Oncogene 15:549-560(1997).
 DR EMBL: AF010404; AAC51735.1; -
 DR InterPro: IPR000910; HMG_12_box.
 DR InterPro: IPR001214; SET.
 DR InterPro: IPR001822; Recombinase.
 DR InterPro: IPR001965; PHD.
 DR InterPro: IPR002965; P_rich_extensn.
 DR InterPro: IPR003616; PostSET.
 DR InterPro: IPR003888; FYRIC_N.
 DR InterPro: IPR003889; FYRIC_C.
 DR Pfam: PF00628; PHD; 3.
 DR Pfam: PF00856; SET; 1.
 DR PRINTS: PR01217; PRICHEXTENS.
 DR PROSITE: PS00398; RECOMBINASES_2; UNKNOWN_1.
 DR PROSITE: PS50280; SET; 2.
 DR SMART: SM00542; FYRC; 1.
 DR SMART: SM00541; FYRN; 1.
 SQ SEQUENCE 5262 AA; 564181 MW; 26B7C7ACAD417E44 CRC64;

DR SMART: SM00398; HMG; 1.
 DR SMART: SM00249; PHD; 4.
 DR SMART: SM00508; PostSET; 1.
 DR SMART: SM00317; SET; 1.
 SQ SEQUENCE 4957 AA; 531848 MW; 1026562E1419CE8D CRC64;

alignment_scores:
 Quality: 60.00 Length: 20
 Ratio: 4.615 Gaps: 0
 Percent Similarity: 65.000 Percent Identity: 45.000

alignment_block:
 US-09-684-016-48411 x 014687 ..

Align seg 1/1 to: 014687 from: 1 to: 4957

14 TGTGAGATGCTTACAACTGATGACCTAATGGCGCAATGGCGGCG 63
 |||::: |||||::: |||||::: |||||::: ||
 930 CYSIALAPROTGYVALGIUGLUSPLEUILLGLINCYSARYHSYCY 946
 64 CACTGACTGG 73
 |||
 946 SGLARGTTP 949

seq_name: sp_human:014686

seq_documentation_block:
 ID 014686; PRELIMINARY; PRT; 5262 AA.
 AC 014686;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE ALR.
 GN ALR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 ON NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97388474; PubMed=9247308;
 RA Prasad R., Zhadanov A.B., Sedkov Y., Bullrich F., Druck T.,
 RA Rallapalli R., Yano T., Alder H., Croce C.M., Huebner K., Mazo A.,
 RA Cnaan E.;
 RT "Structure and expression pattern of human ALR, a novel gene with
 RT strong homology to ALL-1 involved in acute leukemia and to Drosophila
 RT trithorax.";
 RL Oncogene 15:549-560(1997).
 DR EMBL: AF010403; AAC51734.1; -
 DR InterPro: IPR000910; HMG_12_box.
 DR InterPro: IPR001214; SET.
 DR InterPro: IPR001822; Recombinase.
 DR InterPro: IPR001841; Znf_ring.
 DR InterPro: IPR001965; PHD.
 DR InterPro: IPR002965; P_rich_extensn.
 DR InterPro: IPR003616; PostSET.
 DR InterPro: IPR003888; FYRIC_N.
 DR InterPro: IPR003889; FYRIC_C.
 DR Pfam: PF00628; PHD; 5.
 DR Pfam: PF00856; SET; 1.
 DR PRINTS: PR01217; PRICHEXTENS.
 DR PROSITE: PS00398; RECOMBINASES_2; UNKNOWN_1.
 DR PROSITE: PS50280; SET; 2.
 DR SMART: SM00542; FYRC; 1.
 DR SMART: SM00541; FYRN; 1.
 DR SMART: SM00398; HMG; 1.
 DR SMART: SM00249; PHD; 7.
 DR SMART: SM00508; PostSET; 1.
 DR SMART: SM00184; RING; 3.
 DR SMART: SM00317; SET; 1.
 SQ SEQUENCE 5262 AA; 564181 MW; 26B7C7ACAD417E44 CRC64;

alignment_scores:

Quality: 60.00 Length: 20
Ratio: 4.615 Gaps: 0
Percent Similarity: 65.000 Percent Identity: 45.000

alignment_block:

US-09-684-016-48411 x 014686 ..

Align seg 1/1 to: 014686 from: 1 to: 5262

14 TGTGAGATGCTTACGACCTGATGCGCATGCGAGGCTG 63
||||| ||||| ::|||::: |||||::: ||
1255 CysHsAlaProTyValGluGluAspLeuIleGlnCysArgHnScy 1251
64 CACTGACTGG 73
|||
1251 sgluArTrp 1254
seq_name: sp_invertebrate:Q22662

seq_documentation_block:

ID Q22662 PRELIMINARY; PRT; 416 AA.
AC Q22662;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE T22C1.1 PROTEIN.
GN T22C1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA McMurtry A.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Ritken L., Koopra A., Saunders D., Showkeen R.,
RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elegans";
RL Nature 368:32-38(1994).
DR EMBL: Z75550; CAA99920.1; -.
DR InterPro: IPR003126; ZNF_UBR1.
DR Pfam: PF02207; zf-UBR1; 1.
DR SMART: SM00396; ZNF_UBR1; 1.
DR SEQUENCE 416 AA; 47619 MW; 3950A83C4D77E9D2 CRC64;

alignment_scores:

Quality: 59.50 Length: 27
Ratio: 3.719 Gaps: 1
Percent Similarity: 59.259 Percent Identity: 37.037

alignment_block:

US-09-684-016-48411 x Q22662 ..

Align seg 1/1 to: Q22662 from: 1 to: 416

2 AGGTATGTAATGTGAGATGCTTACACCCGTGATGAC.....CT 42
:::||||| ||::: ||| ::|||
197 LysphecysThrCysAspValPheTyrProAspGluAspGlyGlySGI 213

43 AATGGCGCATGCGAGGCTGCACTACTGG 73
::: ||||| ||| |||||

213 uleuGlnCysGlnIleCysGluAspTrp 223

seq_name: sp_plant:Q9T0A1

seq_documentation_block:

ID Q9T0A1 PRELIMINARY; PRT; 443 AA.
AC Q9T0A1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOHETICAL 49.7 KDA PROTEIN.
GN T32A16.30 OR AT4G23860.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Zimmermann W., Gruenisen A., Wambutt R., Bancroft I.,
RA Mewes H.W., Mayer K.F.X., Lemcke K., Schueller C.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Arabidopsis sequencing project;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Zimmermann W., Gruenisen A., Wambutt R., Kalicki J., Wohlmann P.,
RA Smith A., Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL078468; CAB43886.1; -.
DR EMBL: AL161560; CAB81304.1; -.
DR InterPro: IPR001965; PHD.
DR InterPro: IPR003126; ZNF_UBR1.
DR Pfam: PF00628; PHD; 1.
DR Pfam: PF02207; zf-UBR1; 1.
DR SMART: SM00249; PHD; 1.
DR SMART: SM00396; ZNF_UBR1; 1.
KW Hypothetical protein.
SO SEQUENCE 443 AA; 49704 MW; 7DA998CA6ED2904D CRC64;

alignment_scores:

Quality: 59.50 Length: 29
Ratio: 4.250 Gaps: 2
Percent Similarity: 48.276 Percent Identity: 44.828

alignment_block:

US-09-684-016-48411 x Q9T0A1 ..

Align seg 1/1 to: Q9T0A1 from: 1 to: 443

5 TATGTAAATGTGAGATGCTTACACCCGTGATGAC..... 40
||||| ||::: ||||| |||||
128 TyrcysThrCysAspArgProTyr...ProAspProAsnValGluGluGI 143
41CTATGGCGCATGCGAGGCTGCACTACTGG 73
||| ||||| ||| |||||
143 nvalGluMetIleGlnCysCysLeuGluAspTrp 155

seq_name: sp_invertebrate:Q9YRV6

seq_documentation_block:

ID Q9YRV6 PRELIMINARY; PRT; 662 AA.
AC Q9YRV6;

alignment_scores:	Quality: 57.50	Length: 33
	Ratio: 2.212	Gaps: 1
	Percent Similarity: 78.788	Percent Identity: 36.364
alignment_block:		
US-09-684-016-48411/rev x GQ9RV6		..

```

seq_name: sp_fungi:Q9UT79
seq_id: Q9UT79
seq_documentation_block:
  PRELIMINARY:
    PRT: 1588 AA.
    Q9UT79:
      01-MAY-2000 (TREMBLrel. 13, Created)
      01-MAY-2000 (TREMBLrel. 13, Last sequence update)
      01-JUN-2001 (TREMBLrel. 17, Last annotation update)
      PUTATIVE TRANSCRIPTIONAL REGULATORY PROTEIN, PHD FINGER.
      SPAC343.11C.
      Schizosaccharomyces pombe (Fission yeast).
      Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
      Schizosaccharomycetales; Schizosaccharomycetaceae;
      Schizosaccharomycetes.
      NCBI_TaxID=4896;
      [1]
    SEQUENCE FROM N.A.
    STRAIN-972H:
      Murphy L., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
      Submitted (Aug-1997) to the EMBL/GenBank/DBJ databases.
      EMBL: AL109739; CAB52274.1; -
      InterPro: IPR0000345; Cytochrome_bind.
      InterPro: IPR0000561; EGF-like.
      InterPro: IPR001365; PHD.
      InterPro: IPR003347; UMG_C.
      InterPro: IPR003349; UMG_N.
      Pfam: PF00628; PHD; 3.
      Pfam: PF02373; jmjC. 1.
      PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
      PROSITE: PS01186; EGF_2; UNKNOWN_1.
      SMART: SM00545; jmjN; 1.
      SMART: SM00249; PHD; 3.
      SEQUENCE 1588 AA; 180350 MW; 2D18248ABB3DEC55 CRC64;

```

```

alignment_scores:
  Quality: 57.00
  Ratio: 3.562
  Percent Similarity: 72.727
  Percent Identity: 36.364

alignment_block:
  US-09-684-016-48411 x Q9UT79 ..

Align seg 1/1 to: Q9UT79 from: 1 to: 1588

      8 TGTAAATGTGAGATGCCCTTACCAACCGTCGATGACCTAAATGGCCGCAATGCGA 57
          ||| |||:::||||::: |||::: |||::: |||:::
1457 CysLeucysargInlProPhealariIseIseraspIylThryValGInCysH1 1473
      58 GGGCTGCTGACTGACTGG 73
          ::::||| ::::|||
1473 sasncysLeuGIutrp 1478

seq_name: sp_human:Q9BQW0

seq_documentation_block:
  ID Q9BQW0 PRELIMINARY; PRT; 379 AA.
  AC Q9BQW0;
  DT 01-JUN-2001 (Tremblrel. 17, Created)
  DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
  DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
  DE D388517.9.3 (DEATH ASSOCIATED TRANSCRIPTION FACTOR 1 (CONTAINS

```

DE KIAA0333), ISOFORM 3) (FRAGMENT).
 GN DAF1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Smith M.;
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL035669; CAC28883.1; -.
 FT NON_TER 1 1
 SQ SEQUENCE 379 AA; 41245 MW; 3B14597F9243C6D5 CRC64;

alignment_scores:
 Quality: 56.50 Length: 23
 Ratio: 3.531 Gaps: 1
 Percent Similarity: 69.565 Percent Identity: 39.130

alignment_block:
 US-09-684-016-48411 x Q9BQW0 ..

Align seg 1/1 to: Q9BQW0 from: 1 to: 379

```

5 TATGTAAATGTGAGATGCTTACACCGCTGATGACCTAATGGCGCAATG 54
||||| |||:|||||:||||| |||: |||
87 TyrcysllecySarGlnProHisAsn...AsnArgPheMetIleCyScy 102
55 CGAGGCGCTGCACCTGACTGG 73
|:: ||| ::|||
102 sasparGcysgluGluTrrp 108

```

seq_name: sp_human:Q9BQW0

seq_documentation_block:
 ID Q9BQW0 PRELIMINARY; PRT; 544 AA.
 AC Q9BQW0;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
 DE HYPOTHETICAL 59.5 KDA PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY ADENOCARCINOMA;
 RA Strausberg R.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC000770; AAH00770.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 544 AA; 59487 MW; 193341A7ED5FDA CRC64;

alignment_scores:
 Quality: 56.50 Length: 23
 Ratio: 3.531 Gaps: 1
 Percent Similarity: 69.565 Percent Identity: 39.130

alignment_block:
 US-09-684-016-48411 x Q9BQW0 ..

Align seg 1/1 to: Q9BQW0 from: 1 to: 544

```

5 TATGTAAATGTGAGATGCTTACACCGCTGATGACCTAATGGCGCAATG 54
||||| |||:|||||:||||| |||: |||
270 TyrcysllecySarGlnProHisAsn...AsnArgPheMetIleCyScy 285
55 CGAGGCGCTGCACCTGACTGG 73
|:: ||| ::|||
285 sasparGcysgluGluTrrp 291

```

seq_name: sp_human:Q9NUM8

seq_documentation_block:
 ID Q9NUM8 PRELIMINARY; PRT; 562 AA.
 AC Q9NUM8;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
 DE CDNA FL111265 FIS, CLONE PLACE1009158.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE-PLACENTA;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Wagaatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Negahari K., Masuno Y.,
 RA Ninomiya K., Iwayanagi T.;
 RT "NEDO human cDNA sequencing project";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AK002127; BAA92094.1; -.
 DR InterPro: IPR001965; PHD.
 DR Pfam: PF00628; PHD; 1.
 DR SMART: SM00249; PHD; 1.
 SQ SEQUENCE 562 AA; 61425 MW; 3BA79F8DEF9D3D9C CRC64;

alignment_scores:
 Quality: 56.50 Length: 23
 Ratio: 3.531 Gaps: 1
 Percent Similarity: 69.565 Percent Identity: 39.130

alignment_block:
 US-09-684-016-48411 x Q9NUM8 ..

Align seg 1/1 to: Q9NUM8 from: 1 to: 562

```

5 TATGTAAATGTGAGATGCTTACACCGCTGATGACCTAATGGCGCAATG 54
||||| |||:|||||:||||| |||: |||
270 TyrcysllecySarGlnProHisAsn...AsnArgPheMetIleCyScy 285
55 CGAGGCGCTGCACCTGACTGG 73
|:: ||| ::|||
285 sasparGcysgluGluTrrp 291

```

seq_name: sp_human:Q9BRC0

seq_documentation_block:
 ID Q9BRC0 PRELIMINARY; PRT; 562 AA.
 AC Q9BRC0;
 DT 01-JUN-2001 (TREMblrel. 17, Created)
 DT 01-JUN-2001 (TREMblrel. 17, last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
 DE HYPOTHETICAL 61.4 KDA PROTEIN.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE-COLON ADENOCARCINOMA;
 RA Strausberg R.;
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC004237; AAH04237.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 562 AA; 61353 MW; 91A71468F9C5C12 CRC64;

DR InterPro: IPR001209; Ribosomal_S14.
DR InterPro: IPR001965; PHD.
DR InterPro: IPR003618; TFS2_cent.
DR Pfam: PF006628; PHD; 1.
DR PRINTS: PR00163; RUBREDOXIN.
DR PROSITE: PS00527; RIBOSOMAL_S14; UNKNOWN_1.
DR SMART: SM00249; PHD; 1.
DR SMART: SM00510; TFS2M; 1.
FT NON_TER 1
SQ SEQUENCE 991 AA; 107368 MW; 3483D97D3290A7A1 CRC64;

alignment_scores:

Quality: 56.50 Length: 23
Ratio: 3.531 Gaps: 1
Percent Similarity: 69.565 Percent Identity: 39.130

alignment_block:

US-09-684-016-48411 x 015043 ..

Align seg 1/1 to: 015043 from: 1 to: 991

5 TATGTAAATGTGAGATGCGCTTACACCGCTGATGACCTAATGCGCAATG 54
||||| |||:::||||::| |:: |||
36 TyrcysIleCysArgGlnProHisAsn...AsnArgPheMetIleCysCy 51
55 CGAGGGCTGCACCTGACTGCG 73
|::: ||| ::|||
51 SasparGcysGluGluTrp 57

seq_name: sp_human:Q9H4G6

seq_documentation_block:

ID Q9H4G6 PRELIMINARY; PRT; 1042 AA.
AC Q9H4G6;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE DJ88517.9.1 (NOVEL PROTEIN SIMILAR TO MOUSE DEATH INDUCER OBLITERATOR
DE 1 (DIO-1) (CONTAINS KIAA0333) (ISOFORM 1)) (FRAGMENT).
GN DJ88517.9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith M.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL035669; CAC12756.1; -
DR InterPro: IPR001965; PHD.
DR InterPro: IPR001209; Ribosomal_S14.
DR InterPro: IPR003618; TFS2_cent.
DR Pfam: PF006628; PHD; 1.
DR SMART: SM00249; PHD; 1.
DR SMART: SM00510; TFS2M; 1.
DR PROSITE: PS00527; RIBOSOMAL_S14; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 1042 AA; 112916 MW; F1ADB40059F09742 CRC64;

alignment_scores:

Quality: 56.50 Length: 23
Ratio: 3.531 Gaps: 1
Percent Similarity: 69.565 Percent Identity: 39.130

alignment_block:

US-09-684-016-48411 x Q9H4G6 ..

Align seg 1/1 to: Q9H4G6 from: 1 to: 1042

5 TATGTAAATGTGAGATGCGCTTACACCGCTGATGACCTAATGCGCAATG 54
||||| |||:::||||::| |:: |||
::: |||

87 TyrcysIleCysArgGlnProHisAsn...AsnArgPheMetIleCysCy 102
55 CGAGGGCTGCACCTGACTGCG 73
|::: ||| ::|||
102 SasparGcysGluGluTrp 108

seq_name: sp_archaea:Q9YEG2

seq_documentation_block:

ID Q9YEG2 PRELIMINARY; PRT; 544 AA.
AC Q9YEG2;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE HYPOHETICAL 55.8 KDA PROTEIN APE0614.
GN APE0614.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcaceae; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudo Y.,
RA Yamazaki J., Nomura N., Oguchi A., Aoki K.-I., Kudo K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1998).
DR EMBL: AP000060; BAA79584.1; -
DR InterPro: IPR000566; LipidCln_cyLFBP.
DR PROSITE: PS00213; LIPOCALIN; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 544 AA; 55839 MW; 14028D4D558A7891 CRC64;

alignment_scores:

Quality: 56.00 Length: 26
Ratio: 3.111 Gaps: 0
Percent Similarity: 69.231 Percent Identity: 38.462

alignment_block:

US-09-684-016-48411/rev x Q9YEG2 ..

Align seg 1/1 to: Q9YEG2 from: 1 to: 544

116 TGGATTAATTTGAAGATTAAGTATACCTTACATGTGAACCTTACCAGTCA 67
||||: |||::: |||::: |||::: |||
179 TrpIuProIleAspAlaValPheIysValGluIleGluLeuAspArgSe 195
66 GTGCAGCCCTGCGCATTCGCGCATTAGG 39
|:||||| ||||| |||||
195 rCySerProArgIleAlaIalaLeuGly 204

seq_name: sp_bacteria:Q9RLR4

seq_documentation_block:

ID Q9RLR4 PRELIMINARY; PRT; 363 AA.
AC Q9RLR4;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TReMBLrel. 16, Last annotation update)
DE LVH310 PROTEIN.
GN LVH310.
OS Legionella pneumophila.
OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
OC Legionellaceae; Legionella.
OX NCBI_TaxID=446;
RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-PHILADELPHIA-1;
 RX MEDLINE=20032053; PubMed=10564519;
 RA Segal G., Russo J.J., Shuman H.A.;
 RT "Relationships between a new type-IV secretion system and the *lcm/dot*
 RT virulence system of *Legionella pneumophila*.";
 RL Mol. Microbiol. 34:799-809(1999).
 DR EMBL: Y19029; CAB60060.1; ..
 SQ SEQUENCE 363 AA; 39117 MW; F23E77A30A279BC7 CRC64;

alignment_scores:

Quality:	55.50	Length:	37
Ratio:	2.220	Gaps:	1
Percent Similarity:	67.568	Percent Identity:	37.838

alignment_block:

US-09-684-016-48411/rev x Q9RLR4 ..

Align seg 1/1 to: Q9RLR4 from: 1 to: 363

```

122 GTTGTGTGATAAATGAAGATGATCCTTACCTTACATGTAACCTTAC 73
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
241 ValValtRpsasnrglenglnmetThrAsnGlyllelleValThrLeuAs 257
    :
72 CAGTCAGTCAGCCCTCGCATTCGCCCATTAAGTCATCAGGGTTGTAGG 23
    : ||||| ||||| ||||| ||||| ||||| ||||| |||||
257 n.....STProGlySerAspProIleGlyArgSerGlyGlnAlaA 271
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
22 CATCTCACAT 12
    ||||| ||||| ||||| ||||| ||||| ||||| |||||
271 laSPtYrIle 274
  
```

seq_name: sp_human:Q9NR48

seq_documentation_block:

ID Q9NR48 PRELIMINARY; PRT; 2969 AA.

AC Q9NR48:
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE ASH1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20319018; PubMed=10860993;
 RA Nakamura T., Blechman J., Tada S., Rozovskaia T., Itoyama T.,
 RA Bullrich F., Maza A., Croce C.M., Geiger B., Canaan E.;
 RT "hAsh1 protein, a putative transcription factor encoded by a human
 RT homologue of the *Drosophila* ash1 gene, localizes to both nuclei and
 RT cell-cell tight junctions";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:7284-7289(2000).
 DR EMBL: AF257305; AAF68983.1; ..
 DR InterPro: IPR000637; AT_hoek.
 DR InterPro: IPR001025; BAH.
 DR InterPro: IPR001487; Bromodomain.
 DR InterPro: IPR001965; PHD.
 DR InterPro: IPR003616; PostSET.
 DR InterPro: IPR001214; SET.
 DR Pfam: PF01426; BAH; 1.
 DR Pfam: PF00439; bromodomain; 1.
 DR Pfam: PF00628; PHD; 1.
 DR Pfam: PF00856; SET; 1.
 DR SMART: SM00384; AT_hoek; 3.
 DR SMART: SM00439; BAH; 1.
 DR SMART: SM00297; BROMO; 1.
 DR SMART: SM00249; PHD; 1.
 DR SMART: SM00508; PostSET; 1.
 DR SMART: SM00317; SET; 1.
 DR PROSITE: PS50014; BROMODOMAIN_2; 1.

DR PROSITE: PS50280; SET; 1.
 SQ SEQUENCE 2969 AA; 332765 MW; 4696103BD8503493 CRC64;

alignment_scores:

Quality:	55.50	Length:	22
Ratio:	4.269 <td>Gaps:</td> <td>1</td>	Gaps:	1
Percent Similarity:	59.091 <td>Percent Identity:</td> <td>45.455</td>	Percent Identity:	45.455

alignment_block:

US-09-684-016-48411 x Q9NR48 ..

Align seg 1/1 to: Q9NR48 from: 1 to: 2969

```

8 TGTAAATGATGATGCTTACAAACCTGATGACCTAATGGCGAATGCA 57
    ||| ||| :||| ||||| :||| ||||| |||||
2588 CysIleCysGlyLeu...TyraAsnAspGlnGlyLeuMetIleGlnCysAs 2603
    : ||| ||| |||
58 GGGCTGCACCTGACTGG 73
    : ||| ||| |||
2603 pLysCysMetValtRps 2608
  
```

seq_name: sp_human:Q9H0S7

seq_documentation_block:

ID Q9H0S7 PRELIMINARY; PRT; 180 AA.

AC Q9H0S7:
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE HYPOTHETICAL 20.3 KDA PROTEIN.
 GN DKFZP564J1864.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-BRAIN;
 RA Duesterhoft A., Lauber J., Mewes H.W., Gassenhuber J., Wiemann S.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL136660; CAB6595.1; ..
 DR InterPro: IPR000531; TONR_boxC.
 DR PROSITE: PS00430; TONR_DEPENDENT_REC_1; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 180 AA; 20253 MW; DB808E0F3FCDB0 CRC64;

alignment_scores:

Quality:	55.00	Length:	37
Ratio:	2.200 <td>Gaps:</td> <td>0</td>	Gaps:	0
Percent Similarity:	67.568 <td>Percent Identity:</td> <td>32.432</td>	Percent Identity:	32.432

alignment_block:

US-09-684-016-48411 x Q9H0S7 ..

Align seg 1/1 to: Q9H0S7 from: 1 to: 180

```

11 AAATGAGATGGCTTACAAACCTGATGACCTAATGGCGCAATGCGAGG 60
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
60 ArgSerAspLeuGlyPheIleThrSerAspIleThrAlaAspLeuGlnAs 76
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
61 CTGCACCTGACTGTAAGTTTCACATGTGAAGTGATCTACTTATCTCAAT 110
    : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
76 nIlePheAspTrpAsnValIysGlnIleuPheLeuTyrlleuSerAlaGlu 93
    : ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
111 ATTTCACACAA 121
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
93 yIserThrIlys 96
  
```

seq_name: sp_invertebrate:Q9XWD7

seq_documentation_block:

```

ID Q9XWD7 PRELIMINARY; PRT; 368 AA.
AC Q9XWD7;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DE Y47H9C.2 PROTEIN.
GN Y47H9C.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris B.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierrey-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wooldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
DR EMBL: AL032657; CAA21738.1; -.
DR InterPro: IPR001594; Znf-DHHC.
DR Pfam: PF01529; Zf-DHHC; 1.
DR ProDom: PD003041; Znf-DHHC; 1.
SQ SEQUENCE 368 AA; 41180 MW; 10A06804C30E04FA CRC64;

alignment_scores:
Quality: 55.00 Length: 24
Ratio: 3.438 Gaps: 0
Percent Similarity: 66.667 Percent Identity: 45.833

alignment_block:
US-09-684-016-48411/rev x Q9XWD7 ..
Align seg 1/1 to: Q9XWD7 from: 1 to: 368

114 GAATAATGTAAGATAGTATACCTTACATGTGAACCTTACAGTCACTG 65
||| ::::: ||||| |||||::: ||:::
159 GUNHSHValysMetIysTyrCysThrThrcysArgLeuTyrArgProPr 175
64 GCAGCCCTCGCATTCGCCGCAAT 43
::: ||||| ||||| |||||
175 oArgCysSerHisCysAlaIle 182

seq_name: sp.plant:Q9FN15
seq_documentation_block:
ID Q9FN15 PRELIMINARY; PRT; 1516 AA.
AC Q9FN15;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DE GB|AAC80581.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosidis II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-COLOMBIA;

```

```

RX MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Physically assigned of the regions of 1,044,062 bp covered by thirteen
RT DNA Res. 4:291-300(1997).
RL EMBL: AB006699; BAB11682.1; -.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001965; PHD.
DR Pfam: PF00628; PHD; 2.
DR SMART: SM00249; PHD; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN.1.
SQ SEQUENCE 1516 AA; 170012 MW; FFAA8593BA93CEA CRC64;

alignment_scores:
Quality: 55.00 Length: 20
Ratio: 4.583 Gaps: 0
Percent Similarity: 60.000 Percent Identity: 40.000

alignment_block:
US-09-684-016-48411 x Q9FN15 ..
Align seg 1/1 to: Q9FN15 from: 1 to: 1516

14 TGTGAGATGCTTACACCTGATGACTAATGGCGCATGCGAGGCTG 63
||||:||||| ||||| |||||
1198 CysIysIeuProTyrAsnProGlyLeuThrTyrIleHisCysThrSerCy 1214
64 CACTGACTG 73
|||
1214 SASPMETTRP 1217

seq_name: sp.invertebrate:Q9U263
seq_documentation_block:
ID Q9U263 PRELIMINARY; PRT; 1645 AA.
AC Q9U263;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE Y51HA4.12 PROTEIN.
GN Y51HA4.12.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Sulston J.E.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL: AL132952; CAB63382.1; -.
DR InterPro: IPR001214; SET.
DR InterPro: IPR001965; PHD.
DR Pfam: PF00856; SET; 1.
DR Pfam: PF00628; PHD; 1.
DR PROSITE: PS00280; SET; 1.
DR SMART: SM00249; PHD; 1.
DR SMART: SM00317; SET; 1.
SQ SEQUENCE 1645 AA; 181664 MW; 74BF883078BBD57 CRC64;

alignment_scores:
Quality: 54.50 Length: 22
Ratio: 4.192 Gaps: 1

```

Percent Similarity: 59.091 Percent Identity: 40.909

alignment_block:

US-09-684-016-48411 x 09U263 ..

Align seg 1/1 to: 09U263 from: 1 to: 1645

```

      8 TGTAAATGTGAGATGCTTACACACCTGATGACCTAAATGCGCAATGCGA 57
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      797 CysHiscysGlyMetAspHisGlyAspGlyAspThrIle...GlucysG1 812
      58 GGGCTGCACACTGACTGG 73
      ||||| |||
      812 uGlyCysLysThrTrp 817

```

seq_name: sp_invertebrate:044498

seq_documentation_block:

```

ID 044498 PRELIMINARY; PRT: 1655 AA.
AC 044498:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE F15E6.1 PROTEIN.
GN F15E6.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoa; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Altschough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lighting J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Smalton D., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Miller N., Steilyes L., Bradshaw H., Keppler D.;
RL Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (DEC-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF038614; AAB92063.1; -
DR InterPro: IPR001214; SET.
DR InterPro: IPR001965; PHD.
DR Pfam: PF00628; PHD; 1.
DR Pfam: PF00856; SET; 1.
DR PROSITE: PS50280; SET; 1.
DR SMART: SM00249; PHD; 1.
DR SMART: SM00317; SET; 1.
SO SEQUENCE 1655 AA; 182580 MW; 0ABA9D12CB1C0686 CRC64;

```

alignment_scores:

Quality: 54.50 Length: 22
Ratio: 4.192 Gaps: 1
Percent Similarity: 59.091 Percent Identity: 40.909

alignment_block:

US-09-684-016-48411 x 044498 ..

Align seg 1/1 to: 044498 from: 1 to: 1655

```

      8 TGTAAATGTGAGATGCTTACACACCTGATGACCTAAATGCGCAATGCGA 57
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      821 CysHiscysGlyMetAspHisGlyAspGlyAspThrIle...GlucysG1 836
      58 GGGCTGCACACTGACTGG 73
      ||||| |||
      836 uGlyCysLysThrTrp 841

```

seq_name: sp_invertebrate:09VR08

seq_documentation_block:

```

ID 09VR08 PRELIMINARY; PRT: 3680 AA.
AC 09VR08:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CG15637 PROTEIN.
GN DP OR CG15637.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Adyayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,
RA Beeson K.Y., Benton P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borrova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Honck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheel F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao X., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RL "The genome sequence of Drosophila melanogaster."
DR EMBL: AE003576; AAF51000.1; -
DR HSSP: P35555; 1EMN
DR Flybase: FBgn000488; dp.
DR InterPro: IPR000152; Asx_hydroxyl.

```

RA Harris L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegram C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy R., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusserkern D.R., Pacteb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puti V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Schelder F., Shen H.,
RA Rhine B.C., Siden-Kiamos I., Simpkins M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Sytkas R., Teclor C., Turner R., Ventler E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenhach J.,
RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen K.-F., Zavari J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zhong X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
SC Science 287:2185-2195(2000).
DR EMBL: AED03806; AAF57974.1; -;
DR FlyBase, FBgn0034145; CGS005.
SQ SEQUENCE 625 AA; 69682 MW; 8C12CD780F96D5E0 CRC64;

alignment_scores:
Quality: 54.00 Length: 25
Ratio: 3.857 Gaps: 1
Percent Similarity: 56.000 Percent Identity: 36.000

alignment_block:
US-09-684-016-48411 x Q9V7S1 ..

Align seg 1/1 to: Q9V7S1 from: 1 to: 625

5 TATTGTAATGTGAGATG.....CC 24
|||||:|||||||:
270 TYRCYSASNCYSASPARGTHSPVALSERGLUVALLEYRYALAPRPR 286
25 TTACAAACCTGATGCATCATGGG 49
|||||:|||||:|||||:
286 cTyransProaspmsplleileiser 294

seq_name: sp.invertebrate:Q24459

seq_documentation_block:
ID Q24459 PRELIMINARY; PRT; 857 AA.

DC Q24459;
DT 01-NOV-1996 (TrEMBLrel_01, Created)
DT 01-NOV-1996 (TrEMBLrel_01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel_17, last annotation update)
DE POLYCOMBLIKE NUCLEAR PROTEIN.
GN PC1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ISOGENTIC CHROMOSOME 2, DP CN BW;
RX MEDLINE=95044940; PubMed=7936837;
RA Lonie A., Dandrea R., Paro R., Saint R.;
RT "Molecular characterisation of the Polycomblike gene of *Drosophila*
RT melanogaster, a trans-acting negative regulator of homeotic gene
RT expression".;
RL Development 120:2629-2636(1994).
CC -|- SIMILARITY: CONTAINS 2 PHD-FINGER DOMAINS.
DR EMBL: U35153; AAA64457.1; -;
DR FlyBase; FBgn0003044; PCL.
DR InterPro; IPR001965; PHD.
DR InterPro; IPR002999; Tudor.

DR Pfam; PF00628; PHD; 2.
 DR SMART; SM00249; PHD; 2.
 DR SMART; SM00333; TUDOR; 1.
 KW Nuclear protein; DNA-binding.
 FT DOMAIN 242 256 POLY-SER.
 FT DOMAIN 274 277 POLY-THR.
 FT DOMAIN 327 341 POLY-PRO.
 FT DOMAIN 408 411 POLY-GLY.
 FT DOMAIN 424 472 PHD-FINGER.
 FT DOMAIN 512 566 PHD-FINGER.
 FT DOMAIN 843 846 POLY-ARG.
 SO SEQUENCE 857 AA; 95289 MW; D9F0E55797753A2 CRC64;

alignment_scores:
 Quality: 54.00 Length: 23
 Ratio: 4.500 Gaps: 0
 Percent Similarity: 52.174 Percent Identity: 43.478

alignment_block:
 US-09-684-016-48411 x Q24459 ..
 Align seg 1/1 to: Q24459 from: 1 to: 857

5 TATTGTAATGTGAGATGCCTTACAAACCGATGACCTAATGGCGCAATG 54
 ||||| ||| ||| ||| ||| |||||
 514 TTYCYSTYTCYSGLYPSGLYSPHISASmetLeuGlnG 530
 ||| ||| ||| ||| ||| |||
 55 CGAGGCGTCGACGACTGG 73
 ||| ||| ||| ||| ||| |||
 530 sCYsLYsCYsARgAsnTRP 536

seq_name: sp_invertebrate:Q9V8C2

seq_documentation_block:
 ID Q9V8C2 PRELIMINARY: PRT: 1042 AA.

AC Q9V8C2;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PCL PROTEIN.
 GN PCL OR CG5109.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Pfeiffer B.D.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Adil J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferlita S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mewkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheler F., Shen H.,
 RA Shue B.C., Siden Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wesserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RI Science 287:2185-2195(2000).
 RL EMBL: AE003800; AAF57748.1;
 DR FlyBase; FBgn003044; Pcl.
 DR InterPro; IPR001965; PHD.
 DR InterPro; IPR002999; Tudor.
 DR Pfam; PF00628; PHD; 2.
 DR SMART; SM00249; PHD; 2.
 DR SMART; SM00333; TUDOR; 1.
 SO SEQUENCE 1042 AA; 114506 MW; 903323A6C8B6B3B2 CRC64;

alignment_scores:
 Quality: 54.00 Length: 23
 Ratio: 4.500 Gaps: 0
 Percent Similarity: 52.174 Percent Identity: 43.478

alignment_block:
 US-09-684-016-48411 x Q9V8C2 ..
 Align seg 1/1 to: Q9V8C2 from: 1 to: 1042

5 TATTGTAATGTGAGATGCCTTACAAACCGATGACCTAATGGCGCAATG 54
 ||||| ||| ||| ||| ||| |||||
 513 TTYCYSTYTCYSGLYPSGLYSPHISASmetLeuGlnG 529
 ||| ||| ||| ||| ||| |||
 55 CGAGGCGTCGACGACTGG 73
 ||| ||| ||| ||| ||| |||
 529 sCYsLYsCYsARgAsnTRP 535

seq_name: sp_plant:Q9FK65

seq_documentation_block:
 ID Q9FK65 PRELIMINARY: PRT: 570 AA.

AC Q9FK65;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE RECEPTOR-LIKE PROTEIN KINASE.
 DE ARABIDOPSIS THALIANA (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC Eucosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RX MEDLINE=98403884; PubMed=9734815;
 RA Kolani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., Miyajima N.,
 RA Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VI.
 RT Physically assigned P1 and TAC clones.";
 RT DNA Res. 5:203-216(1998).
 RL EMBL; AB012245; BAB09221.1;
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR001611; LRR.
 DR InterPro; IPR003592; LRR_OUT.
 DR InterPro; IPR002290; Set_thr_kin_actsite.

DR InterPro: IPR001245; Tyr_kin.
 DR Pfam: PF00560; LRR; 2.
 DR Pfam: PF00069; pkinase; 1.
 DR PRINTS: PRO0019; LEURKCHRP.
 DR PRINTS: PRO0109; TYRKINASE.
 DR SMART: SM00370; LRR; 4.
 DR SMART: SM00221; STYK; 1.
 DR SMART: SM00220; S_TK; 1.
 DR SMART: SM00219; TYRK; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR ATP-binding; Kinase; Transferase.
 KW SEQUENCE 570 AA; 63537 MW; 972414BDE3A8DBC0 CRC64;

alignment_scores:
 Quality: 53.50 Length: 40
 Ratio: 2.140 Gaps: 3
 Percent Similarity: 62.500 Percent Identity: 40.000

alignment_block:
 us-09-684-016-48411/rev x Q9FG65 ..

Align seg 1/1 to: Q9FG65 from: 1 to: 570

116 TGGATAAATGAGATGATGACCTTACAT.....GT 82
 |||||: :|||||:|||||:|||||: ||
 353 TTPSNARG...ArgIleSerIleAlaIleuGlyAlaAlaGlyLeuVA 368
 ||| :|||:|||||:|||||: |||||:
 81 GAAACTTACACGATGACGACCGCTCGCATTCGCCCATTAAGTCATCAGG 32
 ||| :|||:|||||:|||||: |||||:
 368 IlyLeuHISgluIncysAsnProLysIleIle.....HISArgA 382
 |||||:|||||:|||||:|||||:
 31 GTTGTAGGCACTGCACATT 12
 |||||:|||||:|||||:|||||:
 382 spValLysAlaAlaAsnIle 388

seq_name: sp_invertebrate:Q9VG78

seq_documentation_block:
 ID Q9VG78 PRELIMINARY; PRT; 2016 AA.
 AC Q9VG78;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE CG6525 PROTEIN.
 GN CG6525.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_Taxid=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Franko C., Baldwin D.,
 RA Ballaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokov D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jalali M., Kalush E., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mlishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Teeter C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003695; AAF54807.1;
 DR FlyBase: FBgn0038041; CG6525.
 DR InterPro: IPR001965; PHD.
 DR InterPro: IPR003618; TFS2_cent.
 DR Pfam: PF00628; PHD; 1.
 DR SMART: SM00249; PHD; 1.
 DR SMART: SM00510; TFS2M; 1.
 SQ SEQUENCE 2016 AA; 221998 MW; FB61A332EA710F5E CRC64;

alignment_scores:
 Quality: 53.50 Length: 23
 Ratio: 3.344 Gaps: 1
 Percent Similarity: 69.365 Percent Identity: 39.130

alignment_block:
 us-09-684-016-48411 x Q9VG78 ..

Align seg 1/1 to: Q9VG78 from: 1 to: 2016

5 TATTGTAAATGAGATGACCTTACACCGTATGATGACCGCATG 54
 :|||: |||||:|||||:|||||: :|||: ||
 912 TTPcylIecysArgGlnProHISasn...AsnArgPheMetIleCysCy 927
 |||: ||| |||||
 55 CGAGGCTGCACACTGG 73
 |||: ||| |||||
 927 saspleucysglusaptrp 933

seq_name: sp_virus:Q98VX5

seq_documentation_block:
 ID Q98VX5 PRELIMINARY; PRT; 198 AA.
 AC Q98VX5;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE ENVELOPE PROTEIN (FRAGMENT).
 GN ENV.
 OS Human immunodeficiency virus type 1.
 OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
 OX NCBI_Taxid=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-IND. CAL1058;
 RA Mandal D., Chakrabarti S.;
 RT "Divergent HIV-1 C subtypes isolated from Calcutta, India.";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A3311641; CAC34551.1; -
 FT NON_TER 1
 FT NON_TER 198
 SQ SEQUENCE 198 AA; 22154 MW; 51583AFE0BF1EDC1 CRC64;

alignment_scores: Length: 25
 Quality: 53.00
 Ratio: 2.524
Percent Similarity: 84.000 Percent Identity: 32.000
 Gaps: 0

alignment_block:

US-09-684-016-48411 x Q98VX5 ..

Align seg 1/1 to: Q98VX5 from: 1 to: 198

```
40 CCTAATGGCGCATGGAGGCTGCACCTGACTGTGTAAGTTTCACATGTAA 89
|||||.....|.....|.....|.....|.....|.....|.....|
118 Proasnglythrileargasnleuasnserrilethrilepprocysar 134
90 GGTGACTTATCTTCATTATTTC 114
::: |||.....|.....|.....|.....|.....|.....|.....|
134 gtlelysglnllelleasnmctrip 142
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 19, 2002, 09:32:46 ; Search time 1177.44 Seconds

(without alignments)
1122.546 Million cell updates/sec

Title: US-09-684-016-48411

Perfect score: 123

Sequence: 1 caggtatgttaaatgtaga.....tcaatttaccacacaca 123

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: em_estfun:*
2: em_esthum:*
3: em_estlin:*
4: em_estom:*
5: em_estopl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_hic:*
10: gp_estl:*
11: gp_estl2:*
12: gp_hic:*
13: gp_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_lin:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rod:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	69.4	56.4	530	10	AW761136 s163f12.y
2	67.4	54.8	765	10	BE659470 GM700009B
3	66.2	53.8	764	10	AM349333 GM210007A
4	64.6	52.5	540	11	AM432561 sh75h06.y
5	62.2	50.6	497	11	BC551208 sad34e05.y
6	61.4	49.9	543	10	AM616232 EST307271
7	58.8	47.8	574	10	BE124753 EST393788
8	58.2	47.3	514	10	AM586761 EST18384
9	58.2	47.3	586	11	BE997991 EST429714
10	58.2	47.3	635	10	AM584723 N210838e
11	58.2	47.3	655	10	AM584724 N210839e
12	58.2	47.3	682	10	AM584948 N211110e

13	57.2	46.5	491	11	BG551127 sad33f01.
14	56.6	46.0	335	10	A1486129 EST244450
15	56.6	46.0	608	10	AM035325 EST280888
16	56.6	46.0	633	11	B1208519 EST526559
17	56.6	46.0	732	11	BG887479 EST513330
18	56.4	45.9	541	11	BG048538 OV1_14_H0
19	54.8	44.6	674	10	BE037301 MP19C11 M
20	54.2	44.1	959	10	BE039612 OC01G11 O
21	54	43.9	911	10	BE418763 SCL074_H0
22	53.4	43.4	511	10	A1987319 G60003H05
23	53.4	43.4	675	11	BG605300 WHE2331_B
24	53.4	43.4	726	11	BG643885 EST512079
25	53	43.1	449	11	BG320550 ZMO3_10h1
26	52.8	42.9	672	10	A1731659 BNLGH1105
27	51.8	42.1	194	10	BE123398 B646001F06
28	51.8	42.1	588	10	A1759012 605085F11
29	51.8	42.1	672	10	A1726030 BNLGH1139
30	50.8	41.3	612	10	BE639441 946033B11
31	50.4	41.0	862	11	BF275530 GA_Eb002
32	50.2	40.8	615	11	BF474495 WHE0844_G
33	50.2	40.8	699	10	AU101324 AU101324
34	50.2	40.8	861	11	BG300821 HYSMB001
35	49.6	40.3	420	10	AM459464 sh41e10.y
36	49.4	40.2	539	11	BF650924 NF098E01E
37	49.2	40.0	586	11	BG605406 WHE2329_E
38	49	39.8	418	10	A1443020 sa30h05.y
39	48.6	39.5	380	10	A1026312 L0-529T3
40	48.6	39.5	490	10	AU084649 AU084649
41	48.6	39.5	558	11	BG606610 WHE2958_F
42	48.6	39.5	658	10	AM448025 BRY_1261
43	48	39.0	594	10	BE346433 SP25E05.y
44	48	39.0	706	10	BE214475 HV_CEB000
45	47	38.2	444	10	BE519330 945023A05

ALIGNMENTS

RESULT 1
AW761136
LOCUS s163f12.y1 Gm-cl027 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl027-6024 5' similar to TR:065462 065462 RECEPTOR LIKE PROTEIN
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Glycine max
soybean.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
1 (bases 1 to 530)
AW761136 530 bp mRNA
s163f12.y1 Gm-cl027 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl027-6024 5' similar to TR:065462 065462 RECEPTOR LIKE PROTEIN
; mRNA sequence.

AW761136
AW761136.1 GI:7693038

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Rhanna, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Rheising, B., Allen, M., Bowers, T., Person, B., Sallier, T., Gibbons, M., Pape, D., Harvey, N., Schuck, R., Riller, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or

InfoGenomesystems.com web site: www.genomesystems.com
 Insert Length: 1680 Std Error: 0.00
 High quality sequence stop: 420.
 Location/Qualifiers
 1. 530

/organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl027-6024"
 /clone_1ib="Gm-cl027"
 /tissue_type="cotyledons of 3- and 7-day-old Williams
 seedlings"
 /lab_host="DH10B"

/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
 XhoI; This cDNA library was constructed from mRNA isolated
 from cotyledons of 3- and 7-day-old Williams seedlings
 which were propagated on paper towels with distilled
 water. The cotyledons were flash-frozen in liquid
 nitrogen, then lyophilized for 72 hours. Unequal amounts
 of mRNA was used for cDNA synthesis. Stratagene's cDNA
 Synthesis kit (catalog number 200401) was used to
 synthesize the cDNA. First-stranded synthesis was
 performed with 5-methyl dCTP, hence the ligated cDNA was
 hemimethylated. A modification of Stratagene's
 first-strand synthesis primer was used. An anchor
 nucleotide (V-A, C, or G) was added to the 3' end of the
 primer (GACAGACAGACAGACAGACTGCTCGAG(T)18) to anchor
 the primer at the 5' end of the poly(A) tract. After
 second-strand synthesis, the cDNA ends were filled in
 with cloned Pfu DNA, ligated to EcoRI adapters and
 subsequently phosphorylated. The XhoI site within the
 first-strand synthesis primer was then restricted by
 digestion with XhoI; all XhoI sites in the cDNA would be
 protected by their hemimethylated status. The cDNA
 constructs were size-fractionated with a 500 bp cutoff,
 using GibcoBRL Life Technologies' cDNA Size Fractionation
 column. The column eluent was then ligated into
 Stratagene's pBluescript(II) SK+ Predigested vector
 (pBluescript II SK(+)) that has been digested with EcoRI
 and XhoI, and phosphorylated by Stratagene). 97% of the
 white and blue colonies appear to contain recombinant
 plasmids with cDNA inserts, based on size (n=30). This
 library was constructed by Dr. Paul Keim and Dr. Virginia
 Corryell."

BASE COUNT 136 a 145 c 136 g 113 t
 ORIGIN

Query Match 56.4%; Score 69.4; DB 10; Length 530;
 Best Local Similarity 98.6%; Pred No. 3.3e-12;
 Matches 70; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 gttatgaatgtgagatgcttaacaacctgagcctaattgagcgaatgagggctg 63
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 407 GTATTGTAATGTGAGATGCTTACCAACCTGATGACCTAATGCTGCTCAGGCGCTG 466
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 64 cactgactgctg 74
 ||||||||||||
 DB 467 CACTGACTGCTG 477

RESULT 2
 BE659470/c 765 bp mRNA EST 24-MAY-2001
 LOCUS GM700009B10H8 Gm-r1070 Glycine max cDNA clone Gm-r1070-3447 3',
 DEFINITION
 mRNA sequence.
 ACCESSION BE659470
 VERSION BE659470.1 GI:9985362
 KEYWORDS EST.

ORGANISM soybean.
 Glycine max
 Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
 Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots;
 Rosidae: eustoids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;

REFERENCE 1 (bases 1 to 765)
 AUTHORS Vodka, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Corryell, V.,
 Erpedding, J., Rapp, C., Shoop, E., Pardins, J., Liu, L. and Lewin, H.
 TITLE A Functional Genomics Program for Soybean (NSF 9872565)
 JOURNAL Unpublished (1999)
 COMMENT Other ESTs: AN432561 corresponding to Gm-cl015-5652 (5')
 Contact: Vodka, L.O., PI, A Functional Genomics Program for
 Soybean (NSF 9872565)
 Lewin, H. A., Director, Keck Center for Comparative and Functional
 Genomics
 University of Illinois
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
 Tel: (217) 244-6147
 Fax: (217) 333-4582
 Email: l-vodka@uiuc.edu

This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134. For further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clones@genomesystems.com or InfoGenome
 systems.com web site: www.genomesystems.com
 Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'.
 Location/Qualifiers
 1. 765

FEATURES
 source

/organism="Glycine max"
 /db_xref="taxon:3847"
 /clone="Gm-r1070-3447"
 /clone_1ib="Gm-r1070"

/note="The library Gm-r1070 is a sequence-driven, rerecked
 set of 9,216 clones selected from cDNA libraries from
 various tissues and stages of development of soybean that
 represent 2,639 sequences from immature cotyledons, 1,770
 from immature seed coats, 3,938 from flowers, and 869
 from young pods. The 5' ESTs of the source clones from
 the different libraries was used to select singletons, or
 a representative of each contig, which were rerecked to
 form library Gm-r1070. The cDNA clones of the rerecked
 Gm-r1070 library were then sequenced at the 3' end. The
 contig analysis to select unique genes was performed by
 the laboratory of Ernest Retzel, Center for Computational
 Genomics and Bioinformatics, University of Minnesota,
 http://www.cbc.umn.edu/ResearchProjects/soybean/index.html
 Rerecking was performed by Genome Systems, St. Louis,
 http://www.genomesystems.com, and 3' sequencing by the
 Keck Center for Comparative and Functional Genomics,
 University of Illinois,
 http://www.life.uiuc.edu/biotech/keck.html. Note: The
 corresponding 5' EST from each clone in the Gm-r1070
 library is listed in the 'OTHER EST' field. The detailed
 information on the source library for each clone can also
 be obtained by referring to the Genome Systems clone ID of
 the original cDNA library that is also listed under
 'OTHER EST'."

BASE COUNT 227 a 167 c 150 g 197 t 24 others
 ORIGIN

Query Match 54.8%; Score 67.4; DB 10; Length 765;
 Best Local Similarity 95.8%; Pred. No. 1.6e-11;
 Matches 68; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 gttatgaatgtgagatgcttaacaacctgagcctaattgagcgaatgagggctg 63
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 542 GTATTGTAATGTGAGATGCTTNNAAACCTGATGACCTAATGCTGCAATGCGAGGCTG 483
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 64 cactgactgctg 74
 ||||||||||||
 DB 482 CACTGACTGCTG 472

RESULT 3
 AM349333/c 764 bp mRNA EST 04-OCT-2000
 LOCUS AM349333

DEFINITION GM210007A20A6 Gm-r1021 glycine max cDNA clone Gm-r1021-2316 3', mRNA sequence.

ACCESSION AM349333

VERSION AM349333.1 GI:6847043

KEYWORDS EST.

SOURCE soybean.

ORGANISM Glycine max

REFERENCE 1 (bases 1 to 764)
Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V., Erpelting,J., Rapp,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H. A Functional Genomics Program for Soybean (NSF 9872565)

TITLE Unpublished (1999)

JOURNAL Other-ESTs: At661129

COMMENT Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genome systems.com web site: www.genomesystems.com
Seq primer: 5'-TTTGTGTTTTTTTTTTTTTTT(A/C/G)-3'.

FEATURES
source
1..764
/organism="Glycine max"
/cultivar="Williams"
/db_xref="taxon:3847"
/clone="Gm-r1021-2316"
/clone_lib="Gm-r1021"
/tissue_type="root"
/lab_host="X110-Gold"
/note="Vector: pBluescript II XR, Site_1: EcoRI, Site_2: XhoI; Library Gm-r1021 is a sequence-driven, recloned set of the original library Gm-cl004 which was prepared from root cDNA. The mRNA was isolated from entire roots of 8 day old 'Williams' seedlings which were propagated on paper towels with distilled water. Stragene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. The Gm-cl004 library was constructed by Dr. Paul Keim & Virginia H. Coryell, Department of Biology, Box5640, Northern Arizona University, Flagstaff, AZ 86011, email: paul.keim@na.u.edu, virginia.coryell@na.u.edu. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Computational Biology Centers, University of Minnesota.
http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
Reracking was performed by Genome Systems, St. Louis.
http://www.genomesystems.com and sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois.
http://www.life.uiuc.edu/biotech/keck.html."

BASE COUNT 239 a 161 c 151 g 193 t 20 others

ORIGIN

Query Match 53.8%; Score 66.2; DB 10; Length 764;
Best Local Similarity 95.8%; Pred. No. 3.8e-11;
Matches 68; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 gatttttaaatgtgagatgcttacaacctgctgacctatgagcaatgcaaggctg 63
|||||
Db 566 GTATTGTAATGTGAGATGCTTACCACTGATGACCTAATGTCACATGTCAGGCTG 507

QY 64 cactgactggt 74
|||||
Db 506 CACTGACTGCT 496

RESULT 4
AM432561
LOCUS 540 bp mRNA EST 18-JUL-2000

DEFINITION sh75n06.y1 Gm-cl015 glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl015-5652 5' similar to TR:065462 065462 RECEPTOR LIKE PROTEIN ; mRNA sequence.

ACCESSION AM432561

VERSION AM432561.1 GI:6963868

KEYWORDS EST.

SOURCE soybean.

ORGANISM Glycine max

REFERENCE 1 (bases 1 to 540)
Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna,A., Bolla,B., Matra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Insert length: 1008 Std Error: 0.00
Seq primer: -40RP from Gdbco
High quality sequence stop: 380.

FEATURES
source
1..540
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl015-5652"
/clone_lib="Gm-cl015"
/tissue_type="Mature flowers, field grown plants"
/lab_host="X110-Gold"
/note="Vector: pBluescript II XR, Site_1: EcoRI, Site_2: XhoI; This cDNA library was constructed from mRNA isolated from mature flowers of field grown plants. The cDNA library was prepared using the Stragene pBluescript II XR cDNA library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly (dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into X110-Gold host cells. This library was constructed by Dr. Randy Shoemaker and Dr. John Erpelting."

BASE COUNT 123 a 142 c 138 g 137 t

ORIGIN

Query Match 52.5%; Score 64.6; DB 10; Length 540;
Best Local Similarity 94.4%; Pred. No. 1.2e-10;
Matches 67; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

OY      4  gtattgtaattgagatgagccttaacaacctgatacctaatagagcgaatgcgaggctg 63
          |||||||
DB      468  GTATTGTAATGTGAGATGCTTACCAACCATGATGACTGATGCGAGGCGCTG 527
OY      64  cactgactgct 74
          |||||||
DB      528  CACTGACTGCT 538

RESULT   5
BG551208 497 bp  mRNA EST 09-APR-2001
LOCUS     sa034605.y1 Gm-cl074 glycine max cDNA clone GENOME SYSTEMS CLONE
DEFINITION ID: Gm-cl074-2145 5' similar to TR:09SV14 09SV14 ES43 LIKE PROTEIN
; mRNA sequence.
ACCESSION BG551208
VERSION    BG551208.1 GI:13562988
KEYWORDS   EST.
SOURCE     soybean.
ORGANISM   Glycine max
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
            Glycine.
REFERENCE 1 (bases 1 to 497)
AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
            A., Bolla,B., Marre,M., Hillier,L., Kucaba,T., Martin,L., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schirck
            R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            R., Waterston,R. and Wilson,R.
            Public Soybean EST Project
            Unpublished (1999)
            Contact: Shoemaker R/Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@wustl.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            High quality sequence stop: 380.

FEATURES
    source
        location/Qualifiers
            1..497
            /organism="Glycine max"
            /db_xref="taxon:3847"
            /clone="GENOME SYSTEMS CLONE ID: Gm-cl074-2145"
            /clone_1lb="Gm-cl074"
            /tissue_type="seedlings induced for HR (hypersensitive
            response)"
            /dev_stage="9-11 day old"
            /lab_host="DH10B"
            /note="Vector: pBluescript II SK(+). Site:1: EcoRI; Site:2:
            XhoI. The cDNA library was constructed from mRNA isolated
            from 9-11 day old seedlings that were induced for HR
            (hypersensitive response) by vacuum infiltrating plant
            tissue with Pseudomonas syringae pv. glycinea carrying the
            avrB gene (Genetics 141:1597-1604). Plant tissue (expanded
            unifoliate leaves) was collected at 2, 4, 8, 12, 24, 36,
            and 53 hrs after inoculation and their mRNA pooled equally
            for cDNA construction. The library was prepared using the
            Stratagene pBluescript II SK(+) library construction kit.
            Complementary DNA was synthesized from mRNA using a primer
            consisting of a poly(dT) sequence with an XhoI restriction
            site. EcoRI adaptors were ligated to the blunt-ended cDNA
            fragments followed by XhoI digestion. The cDNA insert is
            protected from XhoI digestion via methylation during first
            strand synthesis. The cDNA fragments were directionally

```

```

BASE COUNT 119 a 140 c 129 g 109 t
ORIGIN

Query Match 50.6%; Score 62.2; DB 11; Length 497;
Best Local Similarity 95.5%; Pred. No. 6.9e-10;
Matches 64; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY      4  gtattgtaattgagatgagccttaacaacctgatacctaatagagcgaatgcgaggctg 63
          |||||||
DB      431  GTATTGTAATGTGAGATGCTTACCAACCATGATGACTGATGCGAGGCGCTG 490
OY      64  cactgactg 70
          |||||||
DB      491  CACTGACTG 497

RESULT   6
AM616232 543 bp  mRNA EST 18-MAY-2001
LOCUS     EST307271 L. hirsutum trichome, Cornell University Lycopersicon
DEFINITION hirsutum cDNA clone CLHT1123 5', mRNA sequence.
ACCESSION AM616232
VERSION    AM616232.1 GI:7322266
KEYWORDS   EST.
SOURCE     Lycopersicon hirsutum.
ORGANISM   Lycopersicon hirsutum.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE 1 (bases 1 to 543)
AUTHORS   van der Hoeven,R.S., Bezzerides,J.L., Matern,A.L., Holt,I.E., Liang
            F., Hansen,T., Craven,M.B., Bowman,C.L., Ronning,C.M., Nierman,W.,
            Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley,S.D.
            Generation of ESTs from wild tomato (Lycopersicon hirsutum)
            trichomes
            Unpublished (2000)
            Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            5 prime sequence.

FEATURES
    source
        location/Qualifiers
            1..543
            /organism="Lycopersicon hirsutum"
            /db_xref="taxon:62890"
            /clone="CLHT1123"
            /clone_1lb="L. hirsutum trichome, Cornell University"
            /tissue_type="trichome"
            /dev_stage="mixed stages"
            /note="Leaves of various stages were shaken in liquid
            nitrogen, shearing off trichomes. This procedure yielded a
            mixture of cells which is highly enriched for trichome
            likely with minor contaminations of other types of leaf
            cells"

BASE COUNT 139 a 134 c 133 g 137 t
ORIGIN

Query Match 49.9%; Score 61.4; DB 10; Length 543;
Best Local Similarity 86.1%; Pred. No. 1.3e-09;
Matches 68; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY      4  gtattgtaattgagatgagccttaacaacctgatacctaatagagcgaatgcgaggctg 63
          |||||||

```

Db	456	GTAACGTAAATGTAAGTGAATGCATACAATCGCATGACCCTCAATGGTTTCATGTGAGGGCTG	515
Qy	64	cactgactggtgaatttca 82 	
Db	516	CAGTGACTGTGTAAGTCTGA 534	
RESULT	7		
LOCUS	BEI24753		
DEFINITION	BEI24753 574 bp mRNA EST -07-SEP-2000 EST9393788 GVN Medicago truncatula cDNA clone pGVN-67F7, mRNA sequence.		
ACCESSION	BEI24753		
VERSION	BEI24753.1	GI:8529310	
KEYWORDS	EST.		
SOURCE	barrel medic.		
ORGANISM	Medicago truncatula		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago. 1 (bases 1 to 574) Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gentl,J.S., Peng, 'H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M. ESTs from one month old nitrogen-fixing root nodules of Medicago truncatula Unpublished (2000)		
AUTHORS	Contact: Carroll P. Vance Department of Agronomy and Plant Genetics University of Minnesota 411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA Tel.: 612 625 5715 Fax: 651-649-5058 Email: vance004@maroon.tc.umn.edu University of Minnesota name:M261621e TIGR sequence name:MFCC28PX More information is available at: http://chrysle.camu.edu/medicago Seq primer: SKmod (CTA GAA CTA gtc gat CC).		
TITLE	Location/Qualifiers		
JOURNAL	1..574		
COMMENT	/organism="Medicago truncatula" /cultivar="genotype A17" /db_xref="taxon:3880" /clone="pgVN-67F7" /clone_1ib="GVN" /tissue_type="N2-fixing root nodules" /dev_stage="effective root nodules harvested one month post inoculation with Sinorhizobium meliloti" /lab_host="E. coli strain XLOLR" /note="Vector: pluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from effective root nodules harvested one month post inoculation with Sinorhizobium meliloti. The cDNA was directionally ligated into the uni-zap XR vector from stragene and packaged using gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-zap phage using Ex-Assist helper phage and propagated in XLOLR cells."		
FEATURES	source		
BASE COUNT	155 a 114 c 137 g 167 t 1 others		
ORIGIN			
Query Match	47.8%; Score 58.8; DB 10; Length 574;		
Best Local Similarity	88.7%; Pred. No. 8.9e-09;		
Matches 63; Conservative 0; Mismatches 8; Indels 0; Gaps 0;			
Qy	4	gtatgtataatgtgagaaccttaaacccgatctaactaatggcgcaatgcagagctg 63 	
Db	439	GTAATTGAATAGTAGAGTCCCTACACCCGATGATCTGATGATTCANTGTGAGGGCTG 498	
Qy	64	cactgactggtgt 74	

DB	499	CAGTCACTGCT	509
RESULT	8		
LOCUS	AM586761	514 bp	mRNA
DEFINITION	ESTJ318384 MHAM Medicago truncatula/Glomus versiforme mixed EST library cDNA clone pMHAM-55G12, mRNA sequence.		
ACCESSION	AM586761		
VERSION	AM586761.1	GI:7266275	
KEYWORDS	EST.		
SOURCE	Medicago truncatula/Glomus versiforme mixed EST library.		
ORGANISM	Medicago truncatula/Glomus versiforme mixed EST library. Eukaryota; mixed EST libraries.		
REFERENCE	1 (bases 1 to 514)		
AUTHORS	Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.		
TITLE	ESTs from roots of Medicago truncatula after colonization with Glomus versiforme		
JOURNAL	Unpublished (2000)		
COMMENT	Contact: Maria J. Harrison Plant Biology Division The Samuel Roberts Noble Foundation 2510 Sam Noble Parkway, Ardmore, OK 73401, USA Tel: 580-223-5810 Fax: 580-221-7380 Email: mjharrison@noble.org Noble EST name:N25126e TIGR sequence name:MTDAV429K More information is available at. 'http://chrstyle.tamu.edu/medicago Seq primer: Skmcd (CTA GAA CTA gtc gat CC). Location/Qualifiers 1. 514 /organism="Medicago truncatula/Glomus versiforme mixed EST library" /cultivar="Medicago truncatula genotype A17" /db_xref="taxon:119092" /clone="pMHAM-55G12" /clone.lib="MHAM" /tissue.type="roots colonized with Glomus versiforme" /dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The library was made from a mixture of RNA from each of these stages." /lab_host="E. coli strain XL0LR" /note="Vector: pBluescript SK-. Site.1: EcoRI; Site.2: XhoI; cDNA was prepared from polyA+ enriched RNA from roots harvested at 10, 17, 22, 31 and 38 days post-inoculation with Glomus versiforme. The cDNA was directionally ligated into the unizap XR vector from stratagene and packaged using gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-zap phage using Ex-assist helper phage and propagated in XL0LR cells."		
FEATURES	Source		
BASE COUNT	135 a	101 c	127 g
ORIGIN	151 t		
Query Match	47.3%	Score 58.2;	DB 10; Length 514;
Best Local Similarity	88.7%;	Pred. No. 1.4e+08;	
Matches 63;	Conservative 0;	Mismatches 8;	Indels 0; Gaps 0;
OY	4	gtatgtgaagtgaagatgccttaaacccgtagtactaataatgagcgaatgagagctg	63
DB	415	gtattgttaaatgtgagatgacctacacccctgattgatctgagtggttcagtgagggctg	474
OY	64	cactgactgctg	74
DB	475	cagtgactgctg	485

REFERENCE 1 (bases 1 to 655)
 AUTHORS Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D.,
 Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
 TITLE ESTs from roots of Medicago truncatula after colonization with
 Glomus versiforme
 JOURNAL Unpublished (2000)
 COMMENT Contact: Harrison M.J.
 Plant Biology Division
 The Samuel Roberts Noble Foundation
 2510 Sam Noble Parkway, Ardmore, OK 73401
 Tel: 580-223-5810
 Fax: 580-221-7380
 Email: mjharrison@noble.org
 Other name: MHAM-7a-P11; Date: 3/14/00; Updated to the Database of
 Expressed Sequence Tags (dbEST) on 04/27/00; More information is
 available at 'http://chrysis.tamu.edu/medicago'.
 Seq primer: T3.
 Location/Qualifiers
 source 1..655
 /organism="Medicago truncatula/Glomus versiforme mixed EST
 library"
 /cultivar="Medicago truncatula genotype A17"
 /db_xref="taxon:119092"
 /clone="MHAM-7K21"
 /clone_1lb="MHAM"
 /tissue_type="roots colonized with Glomus versiforme"
 /dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days
 post-inoculation with Glomus versiforme. The library was
 made from a mixture of RNA from each of these stages."
 /lab_host="E. coli strain XLOLR"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; cDNA was prepared from polyA+ enriched RNA from
 roots harvested at 10, 17, 22, 31 and 38 days
 post-inoculation with Glomus versiforme. The cDNA was
 directionally ligated into the UniZap XR vector from
 Stratagene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-Tap phage using Ex-assist
 helper phage and propagated in XLOLR cells."
 BASE COUNT 207 a 163 c 108 g 177 t
 ORIGIN

Query Match 47.3%; Score 58.2; DB 10; Length 655;
 Best Local Similarity 88.7%; Pred. No. 1.4e-08;
 Matches 63; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 4 gtaatgtaattgagatgacctgacacccctgatactgaatgagcgagagctg 63
 |||||||
 Db 468 gtaattgtaattgagatgacctgacacccctgatactgaatgagcgagctg 409

QY 64 cactgactgct 74
 |||||||
 Db 408 CAGTGACTGCT 398

RESULT 12
 AM584948/c 682 bp mRNA EST 07-SEP-2000
 LOCUS N21110e MHAM Medicago truncatula/Glomus versiforme mixed EST
 DEFINITION library cDNA clone MHAM-7J12, mRNA sequence.
 ACCESSION AM584948
 VERSION AM584948
 KEYWORDS EST.
 SOURCE Medicago truncatula/Glomus versiforme mixed EST library.
 ORGANISM Eukaryota; mixed EST libraries.
 REFERENCE 1 (bases 1 to 682)
 AUTHORS Harrison,M.J., Liu,J., Peng,H., Gonzales,M., Ellis,L., Town,C.D.,
 Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
 TITLE ESTs from roots of Medicago truncatula after colonization with
 Glomus versiforme
 JOURNAL Unpublished (2000)

COMMENT Contact: Harrison M.J.
 Plant Biology Division
 The Samuel Roberts Noble Foundation
 2510 Sam Noble Parkway, Ardmore, OK 73401
 Tel: 580-223-5810
 Fax: 580-221-7380
 Email: mjharrison@noble.org
 Other name: MHAM-7d-E06; Date: 3/14/00; Updated to the Database of
 Expressed Sequence Tags (dbEST) on 04/27/00; More information is
 available at 'http://chrysis.tamu.edu/medicago'.
 Seq primer: T3.
 Location/Qualifiers
 source 1..682
 /organism="Medicago truncatula/Glomus versiforme mixed EST
 library"
 /cultivar="Medicago truncatula genotype A17"
 /db_xref="taxon:119092"
 /clone="MHAM-7J12"
 /clone_1lb="MHAM"
 /tissue_type="roots colonized with Glomus versiforme"
 /dev_stage="Roots harvested at 10, 17, 22, 31 and 38 days
 post-inoculation with Glomus versiforme. The library was
 made from a mixture of RNA from each of these stages."
 /lab_host="E. coli strain XLOLR"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; cDNA was prepared from polyA+ enriched RNA from
 roots harvested at 10, 17, 22, 31 and 38 days
 post-inoculation with Glomus versiforme. The cDNA was
 directionally ligated into the UniZap XR vector from
 Stratagene and packaged using Gigapack III Gold packaging
 extracts. Plasmids containing cDNA inserts were excised
 from the recombinant lambda-Tap phage using Ex-assist
 helper phage and propagated in XLOLR cells."
 BASE COUNT 212 a 172 c 115 g 183 t
 ORIGIN

Query Match 47.3%; Score 58.2; DB 10; Length 682;
 Best Local Similarity 88.7%; Pred. No. 1.4e-08;
 Matches 63; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 4 gtaatgtaattgagatgacctgacacccctgatactgaatgagcgagagctg 63
 |||||||
 Db 468 gtaattgtaattgagatgacctgacacccctgatactgaatgagcgagctg 409

QY 64 cactgactgct 74
 |||||||
 Db 408 CAGTGACTGCT 398

RESULT 13
 BG551127 491 bp mRNA EST 09-APR-2001
 LOCUS sad33f01.y1 Gm-cl074 Glycine max cDNA clone GENOME SYSTEMS CLONE
 DEFINITION ID: Gm-cl074-2162 5' similar to TR:09SVIA 09SVIA ES43 LIKE PROTEIN
 ; mRNA sequence.
 ACCESSION BG551127
 VERSION BG551127.1 GI:13562907
 KEYWORDS EST.
 SOURCE Glycine max
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
 Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 REFERENCE 1 (bases 1 to 491)
 AUTHORS Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Corryell,V., Rhanna
 A., Bolla,B., Matra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
 Wylie,T., Underwood,K., Steptoe,M., Rheising,B., Allen,M., Bowers
 Y., Person,B., Swaller,T., Giddons,M., Pape,D., Harvey,N., Schurk
 R., Rilter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
 R., Waterston,R. and Willson,R.
 TITLE Public Soybean EST Project

JOURNAL
COMMENT

Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
High quality sequence stop: 421.

FEATURES

source
1. 491
Location/Qualifiers
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1074-2162"
/clone_lib="Gm-c1074"
/tissue_type="seedlings induced for HR (hypersensitive response)"
/dev_stage="9-11 day old"
/lab_host="DH10B"
/note="Vector: pBluescript II SK+, Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed from mRNA isolated from 9-11 day old seedlings that were induced for HR (hypersensitive response) by vacuum infiltrating plant tissue with *Pseudomonas syringae* pv. *glyciniae* carrying the *avrB* gene (Genetics 141:1597-1604). Plant tissue (expanded unifoliate leaves) was collected at 2, 4, 8, 12, 24, 36, and 53 hrs after inoculation and their mRNA pooled equally for cDNA construction. The library was prepared using the Stratagene pBluescript II SK(+) library construction kit. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with an XhoI restriction site. EcoRI adaptors were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA insert is protected from XhoI digestion via methylation during first strand synthesis. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into *E. coli* Electromax DH10B host cells. Plant care, inoculations, and library construction were performed by Steve Clough (Lila Vodkin lab, University of Illinois)."
BASE COUNT 117 a 138 c 129 g 107 t
ORIGIN

Query Match 46.5%; Score 57.2; DB 11; Length 491;
Best Local Similarity 95.2%; Pred. No. 2.8e-08;
Matches 59; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 ggtatgtcaatgtgagatgcttaacacctgatacctaagcgcaatgcgaggtc 62
|||||
Db 430 ggtattgtaaatgtgagatgcttaacacctgatacctaagcgcaatgcgaggtc 489
|||||
QY 63 gc 64
||
Db 490 gc 491

RESULT 14
AI486129
LOCUS AI486129 335 bp mRNA EST 18-MAY-2001
DEFINITION EST244450 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
AI486129
ACCESSION AI486129
VERSION AI486129.1 GI:4381500
KEYWORDS EST.
SOURCE tomato.
ORGANISM Lycopersicon esculentum

REFERENCE

AUTHORS
1 (bases 1 to 335)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman,
C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley,
S.D. and Giovannoni,J.
Generation of ESTs from tomato carpel tissue
Unpublished (1999)

JOURNAL

COMMENT
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>.

FEATURES

source
1. 335
Location/Qualifiers
/organism="Lycopersicon esculentum"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="CLEDSK16"
/clone_lib="tomato ovary, TAMU"
/tissue_type="carpel"
/dev_stage="5 days pre-anthesis to 5 days post-anthesis"
/lab_host="XL1-Blue MRF"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2: XhoI; cLED - Tomato Carpel EST Library. Oligo(dT)-primed and directionally cloned cDNA in vector Lambda Zap II with 5' and 3' ends located at the EcoRI and XhoI sites, respectively."
BASE COUNT 99 a 76 c 72 g 88 t
ORIGIN

Query Match 46.0%; Score 56.6; DB 10; Length 335;
Best Local Similarity 87.3%; Pred. No. 4.1e-08;
Matches 62; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 4 gttatgaatgtgagatgcttaacacctgatacctaagcgcaatgcgaggtc 63
|||||
Db 168 gttacttaattgtaaatgtgagatgcttaacacctgatacctaagcgcaatgcgaggtc 227
|||||
QY 64 cactgactgt 74
|||||
Db 228 CACTGACTGCT 238

RESULT 15

AW035325
LOCUS AW035325 608 bp mRNA EST 18-MAY-2001
DEFINITION EST280688 tomato callus, TAMU Lycopersicon esculentum cDNA clone
CLEC40312 similar to A. thaliana ESA3-like protein, mRNA sequence.
AW035325
VERSION AW035325.1 GI:5894081
KEYWORDS EST.

SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asterales; easterids I; Solanales; Solanales; Solanales; Solanaceae; Solanum;
Lycopersicon.

REFERENCE

AUTHORS
1 (bases 1 to 608)
Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.,
Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,
C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
Generation of ESTs from tomato callus tissue
Unpublished (1999)
Contact: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>

TITLE
JOURNAL
COMMENT
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>

[illegible]

```

Db      571 CAGTCACTGTGT 581
|||||
RESULT 18
BG048538
LOCUS
DEFINITION OVI_14_H02.g2_A002 Ovary 1 (OVI) Sorghum bicolor cDNA, mRNA
ACCESSION BG048538
VERSION BG048538.1 GI:12499394
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoidae; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 541)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
,L.H.
TITLE An EST database from Sorghum: ovaries of varying immature stages
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmp@prattuga.edu
Sequences have been trimmed to exclude polyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence
is 20.
Seq primer: PolyTMix
High quality sequence start: 30
High quality sequence stop: 541
POLYA-No.
FEATURES
location/Qualifiers
source 1..541
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Ovary 1 (OVI)"
/note="Orgn: Mix of ovaries of varying immature stages
from 8-week-old plants; Vector: phuescript II from Lambda
zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
from poly-A RNA in the cloning vector Lambda zap II.
Clones to be sequenced were prepared by mass excision."
BASE COUNT 139 a 111 c 141 g 150 t
ORIGIN
Query Match 45.9%; Score 56.4; DB 11; Length 541;
Best Local Similarity 76.7%; Pred. No. 5,3e-08;
Matches 69; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
Yy 4 gatttgtaagtgtgagatgccttaacacctatgacctaaaggcgcatgagggcgtg 63
|||||
Db 52 GTATTGCAAAAGTGTGAGATGCCATACACCCGTGATGACCTCATGGTGCAGCGCTTG 111
|||||
Yy 64 cactgactgctgaagtcttcacatgtaaggtg 93
|||||
Db 112 CAAGGACTGTTCATTCATCTTCATGTG 141
|||||
RESULT 19
BE037301 674 bp mRNA EST 07-JUN-2000
LOCUS MP19C11 MP Mesembryanthemum crystallinum cDNA 5' similar to es43
DEFINITION
ACCESSION BE037301
VERSION BE037301.1 GI:8332317
KEYWORDS EST.
SOURCE common ice plant.
ORGANISM Mesembryanthemum crystallinum

```

```

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Caryophyllidae; Caryophyllales; Aizaceae; Mesembryanthemum.
REFERENCE 1 (bases 1 to 674)
AUTHORS Bohnerl,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea
,H., Kawasaki,S., McColough,A., Michalowski,C.B., Palacio,C.,
Scara,G., Wheeler,M. and Zepeda,G.R.
Functional Genomics of Plant Stress Tolerance
SOURCE Unpublished (2000)
JOURNAL Contact: Michalowski,C.B.
UNIVERSITY OF ARIZONA
Bio Sciences West room 513, Tucson, AZ 85721, USA
Tel: 520-621-7982
Fax: 520-621-1697
Email: cbm@u.arizona.edu.
FEATURES
location/Qualifiers
source 1..674
/organism="Mesembryanthemum crystallinum"
/db_xref="taxon:3544"
/clone_lib="MP"
/tissue_type="apical meristem and leaf primordia"
/dev_stage="6 weeks"
/note="3 d 500mM NaCl"
BASE COUNT 171 a 184 c 164 g 155 t
ORIGIN
Query Match 44.6%; Score 54.8; DB 10; Length 674;
Best Local Similarity 83.8%; Pred. No. 1.8e-07;
Matches 62; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
Yy 1 caggtatgttaagtgtgagatgccttaacacctatgacctaaaggcgcatgaggg 60
|||||
Db 545 CGGTTACTGTAAATGTGAGATGCCCTACATCTGATGACTTGATGAGTTCAATGTAGCGG 604
|||||
Yy 61 ctgactgactgtgt 74
|||||
Db 605 ATGCATGATGTGT 618
|||||
RESULT 20
BE039612 959 bp mRNA EST 07-JUN-2000
LOCUS OC01G11 OC Oryza sativa cDNA 5' similar to receptor like protein,
DEFINITION mRNA sequence.
ACCESSION BE039612
VERSION BE039612.1 GI:8334628
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Euphorbiaceae; Oryzae; Oryza.
REFERENCE 1 (bases 1 to 959)
AUTHORS Bohnerl,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea
,H., Kawasaki,S., McColough,A., Michalowski,C.B., Palacio,C.,
Scara,G., Wheeler,M. and Zepeda,G.R.
Functional Genomics of Plant Stress Tolerance
SOURCE Unpublished (2000)
JOURNAL Contact: Michalowski,C.B.
UNIVERSITY OF ARIZONA
Bio Sciences West room 513, Tucson, AZ 85721, USA
Tel: 520-621-7982
Fax: 520-621-1697
Email: cbm@u.arizona.edu
Insert Length: 1 Std Error: 0.00.
FEATURES
location/Qualifiers
source 1..959
/organism="Oryza sativa"
/strain="pokkali"
/db_xref="taxon:4530"
/clone_lib="OC"
/tissue_type="roots"

```

Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta: Spermatophyta: Magnoliophyta: Liliopsida: Poales: Poaceae: Pooidaeae ; Triticeae; Tritilicum.

REFERENCE	Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
AUTHORS	P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
TITLE	Seaton,C.L., and Tong,J.C.
JOURNAL	The structure and function of the expressed portion of the wheat
COMMENT	genomes - Pre-anthesis spike cDNA library Unpublished (2000)
CONTACT:	Olin Anderson
DEPARTMENT:	US Department of Agriculture, Agriculture Research Service, Pacific
ADDRESS:	West Area, Western Regional Research Center
CITY:	800 Buchanan Street, Albany, CA 94710, USA
TELEPHONE:	Tel: 5105595773
FAX:	Fax: 5105595818
EMAIL:	Email: oanderson@w.usda.gov
SEQUENCE:	Sequence have been trimmed to remove vector sequence and low
PRIMER:	quality sequence with paired score less than 20
SOURCE:	Seq primer: Stratagene SK primer. Location/Qualifiers 1..675
FEATURES	/organism="Triticum aestivum" /cultivar="Chinese Spring" /db_xref="taxon:4565" /clone="WHE2331_B03.D05" /clone_lib="Wheat pre-anthesis spike cDNA library" /tissue_type="Spike before anthesis" /dev_stage="Adult plant" /lab_host="E. coli SOLR" /note="Vector: Lambda Uni-ZAP XR, excised phagemid; Site.1: EcoRI; Site.2: XhoI; Plants were grown in the greenhouse. Whole spike with awns trimmed, white, green and yellow anther were collected and total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised to give Bluescript phagemids in the TJ Close lab (Choi, Close, Fenton) at the University of California, Riverside. Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
BASE COUNT	180 a 172 c 174 g 149 t
ORIGIN	
Query Match	43.4%; Score 53.4; DB 11; Length 675;
Best Local Similarity	79.7%; Pred. No.5.le-07;
Matches	63; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
Dy	4 gtatgtaaatgtgcataccctcaacacctatgactgaaggcacaatgcgaagctg 63 Db 206 GTACGCGAAGTCGCAGATGCCCTAACACCAGATACCTCACTGCTGCATGCGAGGATG 265
Oy	64 cactgcagtgcgttaattca 82 Db 266 CAAAGACTGTGTCATCCA 284
RESULT 24	
LOCUS	BG643885 726 bp mRNA EST 24-APR-2001
DEFINITION	ESTF512079 tomato shoot/meristem Lycopersicon esculentum cDNA clone
ACCESSION	CF0F33K11 5' sequence, mRNA sequence.
VERSION	BG643885
KEYWORDS	BG643885.1 GI:13779110
SOURCE	EST.
ORGANISM	tomato.
SCIENTIFIC NAME	Lycopersicon esculentum
ECOLOGICAL GROUP	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
PHYLUM	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
CLASS	Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
GENUS	Lycopersicon.
STRAIN	1 (bases 1 to 726)
COLLECTOR	van der Hoeven,R., Bezzerides,J., Sun,H., Cho,J., Uterback,T.,
DATE	Hansen,C., Renning,C. and Tanksley,S.
LOCATION	Generation of ESTs from tomato shoot/meristem tissue
REMARKS	Unpublished (2001)

COMMENT	FEATURES	BASE COUNT	ORIGIN
Contact: CUGI Clemson University Genomics Institute Clemson University 100 Jordan Hall, Clemson, SC 29634, USA Email: http://www.genome.clemson.edu/orders/index.html . Location/Qualifiers	source	229 a 148 c 148 g 201 t	<pre> /organism="Lycopersicon esculentum" /cultivar="TA96" /db_xref="taxon:4081" /clone="CTOF3K11" /clone.lib="tomato shoot/meristem" /issue.type="shoot/meristem" /dev_stage="developing shoots from 4-6wks old plants" /lab_host="SOLR" /xmol="Vector: pBluescript SK(-); Site_1: ECORI; Site_2: XhoI; Small expanding leaves from the growing tip were taken from greenhouse plants (4-6wks old TA96). Tissue was immediately frozen in liquid nitrogen." </pre>
Query Match Best Local Similarity 84.5%; Pred. No. 5.2e-07; Matches 60; Conservative 0; Mismatches 11; Indels 0; Gaps 0;			
QY 4 gttatgtaatgtgagatgccttaacaacctgattaccataaggcgagatcgagagctg 63 Db 43 GTATGTGAATGTGGAATGCCCTACACCCGTGATGATCTCATGTGTACAAATGTGAAGATG 102 			
QY 64 cactgactgagt 74 			
Db 103 CAAGAAGTGTGT 113			
RESULT 25 BG320550/c			
LOCUS DEFINITION ACCESSION VERSION KEYWORDS SOURCE ORGANISM	<pre> BG320550 449 bp mRNA EST 27-FEB-2001 Zm03.10h12.A Zm03.AAFC.ECORC.cold_stressed_maize_seedlings Zea mays CDNA clone Zm03.10h12, mRNA sequence. BG320550 BG320550.1 GI:13150228 EST. Zea mays. Zea mays. Zea mays. Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea. 1 (bases 1 to 449) Simmonds,J.A., Singh,J.A., Piche,C., Cass,L., Couroux,P., De Moors A., Harris,L.J., Hattori,J.I., Ouellet,T., Robert,L.S., Sprott,D. and Tinker,N.A. </pre>		
REFERENCE AUTHORS	<p>Expressed Sequence Tags from Cold-Stressed Maize Seedlings Grown Under High Light Intensity Unpublished (2001) Contact: Singh,J.A. Eastern Cereal and Oilseed Research Centre Agriculture and Agri-food Canada 960 Carling Avenue, Bldg. 20, Ottawa, Ontario, K1A 0C6, Canada Tel: (613) 759-1862 Fax: (613) 759-1701 Email: singhja@em.agr.ca. Location/Qualifiers</p>		
TITLE JOURNAL COMMENT	<pre> 1. 449 /organism="Zea mays" /cultivar="CO328" /db_xref="taxon:4577" /clone.lib="Zm03.10h12" /clone.lib="Zm03.AAFC.ECORC.cold_stressed_maize_seedlings" /issue.type="leaf, crown" /dev_stage="4-leaf" /note="Vector: Bluescript SK-/XhoI-ECORI; Site_1: Eco RI; </pre>		
FEATURES source			

ACCESSION A1755912
VERSION A1755912.1 GI:5152847
KEYWORDS EST.
SOURCE zea mays.
ORGANISM zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE	1 (bases 1 to 588)
AUTHORS	Walbot,V
TITLE	Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL	University
COMMENT	Unpublished (1999)
	Contact: Walbot V

FEATURES	Location/Qualifiers
source	1. .588

BASE COUNT	161 a	163 c	122 g	142 t
ORIGIN				

Query Match	42.1%	Score 51.8	DB 10	Length 588
Best Local Similarity	78.5%	Pred. No. 1.6e-06		
Matches 62; Conservative	0;	Mismatches 17;	Indels 0;	Gaps 0;

Oy	4	gtatctaaagttagatgccttaacacccgtgacttaatgcgcaatgcgaggtc	63
Db	539	gtattgcgaatgtagatgccctgcaacccgagtagactctatgctgcagttccatgcttg	480
Oy	64	cactgcactggtgaagtctca	82
Db	479	caagcactggtggtccatcca	461

RESULT	29
AI726030	
LOCUS	
DEFINITION	
AI726030	672 bp mRNA EST 11-JUN-1999
BNHG113381	six-day cotton fiber <i>Gossypium hirsutum</i> cDNA 5' similar to (AI022140) receptor like protein (fragment) [<i>Arabidopsis thaliana</i>], mRNA sequence.

REFERENCE	1 (bases 1 to 672)
AUTHORS	Blewitt, M., Matz, E. C., Davy, D. F. and Burr, B
TITLE	Ests from developing cotton fiber
JOURNAL	Unpublished (1999)
COMMENT	Contact: Ben Burr

```
FEATURES
  source      location/Qualifiers
1. .672
/organism="Gossypium hirsutum"
/cultivar="Acala Maxxa"
/db_xref="taxon:3635"
```

BASE COUNT	ORIGIN
174 a	162 c 145 g 190 t 1 others

Query Match	42.1%;	Score 51.8;	DB 10;	Length 672;
Best Local Similarity	78.5%;	Pred. No. 1.7e-06;		
Matches 62;	Conservative 0;	Mismatches 17;	Indels 0;	Gaps 0;

QY	4	gltitlaaibtytgagaaagccttcaaacccctgaaacctaatgycgaatgtggaagtgctg	63
Db	592	gTATTCAATGTGAATGCCATATTAACCCGAGTACCTATGTGTCATCGGAAGATG	651
QY	64	caactactgttaagtcca	82
Db	652	CAAGAGATTGTTTCATCCA	670

RESULT	30
BE639441	
LOCUS	BE639441 612 bp mRNA
DEFINITION	94603JB11.y2 946 - tassell primordium prepared by Schmidt lab ze
	mays cDNA, mRNA sequence.

REFERENCE AUTHORS TITLE	JOURNAL
1 (bases 1 to 612) Walbot, V. Maize ESTs from various cDNA libraries sequenced at Stanford University Unpublished (1999)	
Contact: Walbot V.	

FEATURES	Location/Qualifiers
source	1. .612

BASE COUNT	134 a	176 c	186 g	116 t
ORIGIN				

Query Match	41.3%;	Score 50.8;	DB 10;	Length 612;
Best Local Similarity	82.9%;	Pred. No. 3.5e-06;		
Matches 58;	Conservative	0;	Mismatches 12;	Indels 0;
				Gaps 0;

QY 5 tattgtaaatgtgagatgccttacaacccgtgatgacctaatggcgcaatgcgagggtgc 64

```

Db      463  TATTGCAAGTGTGAGATGCCGTACAAACCCGATGACCTCATGTGCAAGTATGC 522
OY      65  actgactgt 74
Db      523  AAGGACTGGT 532

RESULT 31
LOCUS    BF275530
DEFINITION  GA_Eb0024E17f Gossypium arboreum 7-10 dpa fiber library Gossypium
ACCESSION BF275530
VERSION    BF275530.1 GI:112066600
KEYWORDS   EST
SOURCE     Gossypium arboreum.
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE  1 (bases 1 to 862)
AUTHORS    Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
            D., Wood,T.C., Leslie,A. and Wilkins,T.A.
TITLE      An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
JOURNAL    Unpublished (2000)
COMMENT    Contact: Wing RA
            Clemson University Genomics Institute
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TATACGACTCACTATAGGG
            High quality sequence start: 80.
            High quality sequence stop: 680.
FEATURES   Location/Qualifiers
            source          1..862
                        /organism="Gossypium arboreum"
                        /strain="AKA"
                        /cultivar="8400"
                        /db_xref="taxon:29729"
                        /clone="GA_Eb0024E17f"
                        /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                        /tissue_type="Fibers isolated from bolls harvested 7-10
                        dpa"
                        /lab_host="E. coli"
                        /note="Vector: PBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 223 a 213 c 214 g 211 t 1 others
ORIGIN

Query Match          41.0%; Score 50.4; DB 11; Length 862;
Best Local Similarity 71.7%; Pred. No. 5e-06;
Matches 66; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

OY      4  gttatgtaattgagatgacctaccacccctgatactgacgaatgacgaaggctg 63
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      597  GTATTGCAAGTGTGAGATGCCGTACAAACCCGATGACCTCATGTGCAAGTATGC 656
OY      64  cactgactgtgtaagtgtacatgtaagtgtat 95
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      657  CAAGATCGGTTTCCTCATCTGCATGAGAT 688

RESULT 32
LOCUS    BF474495
DEFINITION  BF474495 615 bp mRNA EST 04-DEC-2000
ACCESSION WHE0844.G02_N04ZS WHEAT vernalized crown cDNA library Trilicium
            aestivum cDNA clone WHE0844.G02_N04, mRNA sequence.
VERSION    BF474495
            BF474495.1 GI:11543677

```

```

KEYWORDS  EST.
SOURCE    bread wheat.
ORGANISM  Trilicium aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triliceae; Trilicium.
REFERENCE  1 (bases 1 to 615)
AUTHORS    Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
            P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
            Seaton,C.L. and Tong,J.C.
TITLE      The structure and function of the expressed portion of the wheat
            genomes - Vernalized crown cDNA library
JOURNAL    Unpublished (2000)
COMMENT    Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oanderson@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES   Location/Qualifiers
            source          1..615
                        /organism="Trilicium aestivum"
                        /cultivar="Chinese Spring"
                        /db_xref="taxon:4565"
                        /clone="WHE0844.G02_N04"
                        /clone_lib="Wheat vernalized crown cDNA library"
                        /tissue_type="Crown tissue of seedling"
                        /dev_stage="Five-week old seedling"
                        /lab_host="E. coli SOLR"
                        /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                        Site_1: EcoRI; Site_2: XhoI; Seeds were germinated and
                        grown at 4 C for 5 weeks. The tissue, total RNA, and
                        poly(A) RNA were prepared. A cDNA library was made, and
                        the cDNA clones were in vivo excised to give phuscript
                        phagemids in the T7 close lab (Choi, Close, Fenton) at
                        the University of California, Riverside. Plasmid DNA
                        preparations and DNA sequencing were performed in the OD
                        Anderson lab (all other authors)."
BASE COUNT 175 a 120 c 154 g 166 t
ORIGIN

Query Match          40.8%; Score 50.2; DB 11; Length 615;
Best Local Similarity 81.7%; Pred. No. 5.5e-06;
Matches 58; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

OY      4  gttatgtaattgagatgacctaccacccctgatactgacgaatgacgaaggctg 63
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      173  GTTTGCAAGTGTGAGATGCCGTACAAACCCGATGACCTCATGTGCAAGTATGC 232
OY      64  cactgactgt 74
            ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      233  CTCTGACTGGT 243

RESULT 33
LOCUS    AU101324
DEFINITION  AU101324 Rice panicle at ripening stage Oryza sativa cDNA clone
ACCESSION AU101324
VERSION    AU101324.1 GI:9865574
KEYWORDS   EST
SOURCE     Oryza sativa.
ORGANISM   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzae; Oryza.
REFERENCE  1 (bases 1 to 699)

```



```

/tissue_type="vegetable buds of field grown plants"
/lab_host="XL10-Gold"
/Note="Vector: Bluescript II Xr, Site 1: EcoRI; Site 2:
XhoI; This cDNA library was constructed from mRNA isolated
from vegetable buds of field grown plants. The cDNA
library was prepared using the Stratagene Bluescript II
XR library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dG) sequence with a XhoI restriction site. EcoRI adaptors
were ligated to the blunt-ended cDNA fragments followed by
XhoI digestion. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
Bluescript vector. The ligated cDNA fragments were
transformed into XL10-Gold host cells. This library was
constructed by Dr. Randy Shoemaker and Dr. John
Erpelting."

BASE COUNT      111 a      74 c      114 g      121 t
ORIGIN

Query Match      40.3%; Score 49.6; DB 10; Length 420;
Best Local Similarity 80.6%; Pred. No. 7.9e-06;
Matches 58; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY      4      gttatgtaatgtgagatgcttaccacacctgatactgaatgagcgaaggctg 63
         ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      210      GTATTGCAAGTGGAGATGGCTTATAACCAAGATGACCTGATGTGTCAGAGTGAAGGGGTG 269

QY      64      cactgactgta 75
         || | | | | | |
Db      270      CAAGGATTGGTA 281

RESULT 36
LOCUS      BF650924      539 bp      mRNA      EST      20-DEC-2000
DEFINITION      NF098E01EC1F1005 Elicited cell culture Medicago truncatula cDNA
ACCESSION      BF650924
VERSION      BF650924.1 GI:11916054
KEYWORDS      EST.
SOURCE      barrel medic.
ORGANISM      Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
1 (bases 1 to 539)
Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
Flores,H.R., Imman,J.T., Weller,J.W. and May,G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation -
Center for Medicago genomics Research
Unpublished (2000)
Contact: Dixon RA
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7302
Fax: 580 221 7380
Email: radixon@noble.org
Insert Length: 539 Std Error: 0.00
Plate: 098 row: E column: 01
Seq primer: TCACACAGGAACACGCTATGAC.

FEATURES
source
1..539
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF098E01EC"
/tissue_type="Elicited cell culture"
/dev_stage="Cell suspensions derived from root tissues"
/Note="Cells were induced six days after subculture"
/Note="Vector: Lambda Zap; Cells were induced with yeast

```

```

cell wall extracts equivalent to 50ug/ml glucose in the
final concentration. Samples were taken at 0.5, 1, 12 and
24 hours after induction. Equal amounts of RNA from each
time point were pooled and used for mRNA isolation."

BASE COUNT      146 a      108 c      125 g      157 t      3 others
ORIGIN

Query Match      40.2%; Score 49.4; DB 11; Length 539;
Best Local Similarity 88.9%; Pred. No. 9.7e-06;
Matches 64; Conservative 0; Mismatches 7; Indels 1; Gaps 1;

QY      4      gttatgtaatgtgagatgcttaccacacctgatactgaatgagcgaatgagcgaaggct 62
         ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      443      GTATTGTAATGTGAGATGGCTTACCAACCTGATGATGTGTTCAANTGTAGGGCT 502

QY      63      gcaactgactgct 74
         || | | | | | |
Db      503      GCANTGACTGCT 514

RESULT 37
LOCUS      BG605406      586 bp      mRNA      EST      16-APR-2001
DEFINITION      WHE2329_E06_11Z5 Wheat pre-anthesis spike cDNA library Trifolium
aestivum cDNA clone WHE2329_E06_111, mRNA sequence.
ACCESSION      BG605406
VERSION      BG605406.1 GI:13635409
KEYWORDS      EST.
SOURCE      bread wheat.
ORGANISM      Trifolium aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
Triticeae; Triticum.
1 (bases 1 to 586)
Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
Seaton,C.L. and Tong,J.C.
The structure and function of the expressed portion of the wheat
genomes - Pre-anthesis spike cDNA library
Unpublished (2000)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 51059595773
Fax: 51059595818
Email: oanderson@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low
quality sequence with phred score less than 20
Seq primer: Stratagene SK primer.

FEATURES
source
1..586
/organism="Trifolium aestivum"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="WHE2329_E06_111"
/clone_11b="Wheat pre-anthesis spike cDNA library"
/tissue_type="Spike before anthesis"
/dev_stage="Adult plant"
/lab_host="E. coli SOLR"
/Note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Plants were grown in the
greenhouse. Whole spike with awns trimmed, white, green
and yellow anther were collected and total RNA, and
poly(A) RNA were prepared, a cDNA library was made, and
the cDNA clones were in vivo excised to give bluescript
phagemids in the T7 close lab (Choi, Close, Fenton) at
the University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (all other authors)."

BASE COUNT      149 a      157 c      137 g      142 t      1 others
ORIGIN

```

BASE COUNT	105 a	70 c	115 g	128 t
ORIGIN	vargina.coryell@neu.edu			
Query Match	39.8%	Score 49;	DB 10;	Length 418;
Best Local Similarity	79.5%	Pred. No. 1.2e-05;		
Matches	58; Conservative	0; Mismatches	15; Indels	0; Gaps
OY	4	gtattgaatagtgtgagatgccttacaccctgatagacctaaatgtgcgcgaatgcgagagctg	63	
Db	212	GTATTGCAAGTGCAGCGAGCTTATTAACCCAGATGACCTCATGGGCGACGTGGAAGGCTG	271	
OY	64	cactgcactgctaa	76	
Db	272	CAAGGATTGANA	284	
RESULT	39			
LOCUS	AI026312	380 bp	mRNA	EST
DEFINITION	L0-529T3 Ice plant Lambda Uni-Zap XR expression library. 0 hours			20-FEB-2001
ACCESSION	AI026312			
VERSION	AI026312.1	GI:3244138		
KEYWORDS	EST.			
ORGANISM	common ice plant.			
SOURCE	Mesembryanthemum crystallinum			
REFERENCE	1 (bases 1 to 380)			
AUTHORS	Cushman, J.C.			
TITLE	An expressed sequence tag database for the common ice plant,			
JOURNAL	Mesembryanthemum crystallinum			
COMMENT	Unpublished (1997)			
	Contact: Cushman JC			
	Department of Biochemistry			
	University of Nevada			
	MS200, Reno, NV 89557-0014, USA			
	Tel: 775-784-1918			
	Fax: 775-784-1650			
	Email: jcushman@unr.edu			
	PCR PRIMERS			
	FORWARD: T7			
	BACKWARD: T3			
	Plate: L0-6	row: D	column: 1	
	Seq primer: T3			
	High quality sequence stop: 340.			
	Location/Qualifiers			
	1. 380			
	/organism="Mesembryanthemum crystallinum"			
	/db_xref="taxon:3544"			
	/clone="L0-529"			
	/clone_lib="Ice plant Lambda Uni-Zap XR expression library			
FEATURES				
source				

RESULT	42		
AW448025			
LOCUS	AW448025	658 bp	mrna
			EST
			03-JAN-2001

Fax: 314 286 1810
Email: estwatson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com

High quality sequence stop: 449.

Location/Qualifiers

1..594

/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1042-1593"
/clone.lib="Gm-c1042"
/tissue.type="Whole seedling without cotyledons"
/lab.host="DH10B"
/note="vector: plusscriptII SK+, site_1: EcoRI, site_2: XhoI; This cDNA library was constructed from mRNA isolated from 2 week old seedlings with the cotyledons removed at the time of harvest. The seedlings for the cultivar Raiden were grown in a growth chamber using germination paper. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pluscript vector. The ligated cDNA fragments were transformed into DH10B host cells (Gibco BRL). This library was constructed by Dr. Randy Shoemaker."

Fax: 864 656 4293
Email: rtwing@clermson.edu
Seq primer: AATTACCTCCTAAGGCG
High quality sequence start: 4
High quality sequence stop: 630.

FEATURES

source

```

1. 706
/organism="Hordeum vulgare"
/cultivar="C116151 (M1a6)"
/db_xref="taxon:4513"
/clone="HV_CEB000301f"
/library="HVCNDA0005 (Erysiphe infected & control)"
/tissue_type="seedling green leaf"
/lab_host="SOLR"
/notes="vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI; For
more details on library preparation and sequence analysis
see http://www.genome.clemson.edu/projects/barley/ To
order a clone see http://www.genome.clemson.edu/orders"

BASE COUNT      190 a      182 c      147 g      187 t
ORIGIN

```

Query Match

Best Local Similarity 39.0%; Score 48; DB 10; Length 706;
Matches 60; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

```

QY  4  gttcttaatgtgagatgcttacaacctgtgacctaatgagcgcaatgcgaggtcg 63
    || || || || || || || || || || || || || || || || || || || ||
DB  605  GTACTCGAAGTGGCATGCTGTACCAACCGAGATGCTGATGTCAGTGTGAGGATG 546
    || || || || || || || || || || || || || || || || || || || ||
QY  64  cactgactgtgaattcac 83
    || || || || || || || || || || || || || || || || || || || ||
DB  545  CAAAGCCTGCTTCATCCAC 526

```

RESULT 45

BE519330

LOCUS

DEFINITION BE519330 444 bp mRNA EST 08-AUG-2000
945023A05.y1 945 - Mixed adult tissues from Walbot lab, same as 707

ACCESSION BE519330 (SK) Zea mays cDNA, mRNA sequence.

VERSION BE519330

KEYWORDS EST.

SOURCE Zea mays.

ORGANISM Zea mays.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 444)

REFERENCE 1 (bases 1 to 444)
Walbot V.
Maize ESTs from various cDNA libraries sequenced at Stanford

AUTHORS University

TITLE Unpublished (1999)

JOURNAL Contact: Walbot V

COMMENT Department of Biological Sciences

Stanford University
855 California Ave, Palo Alto, CA 94304, USA

Tel: 650 723 2227

Fax: 650 725 8221

Email: walbot@stanford.edu

Plate: 945023 row: A column: 05.

FEATURES Location/Qualifiers

1..444

/organism="Zea mays"

/cultivar="W23"

/db_xref="taxon:4577"

/clone_lib="945 - Mixed adult tissues from Walbot lab,

same as 707 (SK)"

/tissue_type="tassel, kernal, silk, husk, root, leaf"

/dev_stage="fully-grown"

/lab_host="DH10B"

/note="Organ: tassel, kernal, silk, husk, root, leaf;

Vector: pGAD10; Site_1: EcoRI; cDNA library from fully
differentiated maize tissues from an active Mutator plant.
Tissue ratio is 4:2:1:1:1 (tassel, kernal, silk, husk,
root, leaf). Unidirectionally cloned. New library number
given to library 707 for additional sequencing."

```

BASE COUNT      124 a      95 c      109 g      116 t
ORIGIN

```

Query Match

Best Local Similarity 38.2%; Score 47; DB 10; Length 444;
Matches 56; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

```

QY  4  gttcttaatgtgagatgcttacaacctgtgacctaatgagcgcaatgcgaggtcg 63
    || || || || || || || || || || || || || || || || || || || ||
DB  237  GTTTGCAAGTGGCATGCTGTACCAACCGAGATGCTGATGTCAGTGTGAGGATG 296
    || || || || || || || || || || || || || || || || || || || ||
QY  64  cactgactgtgt 74
    || || || || || || || || || || || || || || || || || || || ||
DB  297  TTCTGACTGCT 307

```

Search completed: March 19, 2002, 10:18:40
Job time: 2754 sec

PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150984.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151308.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160747.
PR 21-OCT-1999; 99US-0160761.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores: Quality: 138.00 Length: 23
 Ratio: 6.273 Gaps: 0
Percent Similarity: 95.652 Percent Identity: 91.304

alignment_block:
US-09-684-016-48411 x AG05461 ..
Align seg 1/1 to: AG05461 from: 1 to: 186


```
5 TATGTAAATGTGAGATGCTTACACCGCTGATGACCTAATGGCAATG 54
|||||
110 TtCysLysCysLysgluMetProTyrAsnProkspAspleuMetValGlnC 126
55 CGAGGCGCTGCACCTGACTGG 73
|||||
126 sGLUGLYCysLysAspTrp 132

seq_name: /STD57/gcgdata/geneseq/geneseq/AA2000.DAT:AA06765

seq_documentation_block:
ID AAC06765 standard; Protein: 186 AA.
XX
AC AAC06765;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 3659.
XX
KM Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0133256.
PR 14-MAY-1999; 99US-01334218.
PR 14-MAY-1999; 99US-01334219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 18-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145318.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145351.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
```

```

PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148564.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

alignment_scores:
  Quality: 138.00
  Ratio: 6.273
  Gaps: 23
  Percent Similarity: 95.652
  Percent Identity: 91.304
alignment_block:

```

```

US-09-684-016-48411 x AAC06765 ..
Align seq 1/1 to: AAC06765 from: 1 to: 186
5 TATGTAAATGTGATGCCTTCAACCTGATGACCTATGCGGAATG 54
|||||
110 TTYCTSLYSCYSGLUKCTPRYYRASNPROSPASPLEUETVALGNCY 126
55 CGAGGCGCTGCACCTGACTGTG 73
|||||
126 SGLUGLYCYSLYSASPTIR 132
seq_name: /SIDS2/gcgsdata/geneseq/geneseq/AA2000.DAT:AAG36118
seq_documentation_block:
ID AAC36118 standard; Protein; 186 AA.
XX
AC AAG36118;
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 44217.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX
XX Zea mays subsp. mays.
XX
XX EPI033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131445.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134376.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137522.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.

```

PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141642.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0152363.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154479.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:

Quality: 138.00
Ratio: 6.273

Length: 23

Gaps: 0

Percent Similarity: 95.652 Percent Identity: 91.304

alignment_block:

US-09-684-016-48411 x AAG36118 ..

Align seg 1/1 to: AAG36118 from: 1 to: 186

```
5 TATGTGAATGTGATGCTTACACACCGATGACCTAATGGCGCAATG 54
|||||
110 TycCylscYscYsclumetProTYRasnProaspspleumetValGlnCY 126
55 CGAGGCTGCACCTGACTGG 73
|||||
126 sGIuGIcYcYsLYsAsPTrP 132
```

seq_name: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT:AAG39943

seq_documentation_block:

ID AAG39943 standard; Protein: 186 AA.

AC AAG39943;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 49493.

KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

OS Arabidopsis thaliana.

XX EPI033405-A2.

PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 23-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.

```

alignment_scores:
  Quality: 138.00
  Ratio: 6.273
  Percent Similarity: 95.652
  Percent Identity: 91.304

alignment_block:
  US-09-684-016-48411 x AAG39943
  ..

Align seg 1/1 to: AAG39943 from: 1 to: 186

5 TTTTGTGAATGTGAGATGCGCTTACAAACCGTATGACGACCAATGGCGCAATG 54
|||||
110 TTYCYLVSCLYSCGLMETPTOTRYNSPPOASPSAPLEUETVALGNCY 126
55 CGAGGCGCTGCACCTGACTGG 73
|||||
126 scluuglycylsaspttrp 132

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT:AAG39942

seq_documentation_block:
ID AAG39942 standard; Protein: 191 AA.
XX
AC AAG39942;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thailana protein fragment SEQ ID NO: 49492.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hydridisation assay; genetic mapping; gene expression control; promoter
KW termination sequence.
XX
OS Arabidopsis thailana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
FE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.

```

PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142053.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144864.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145912.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157173.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161922.

PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:

Quality: 138.00 Length: 23
Ratio: 6.273 Gaps: 0
Percent Similarity: 95.652 Percent Identity: 91.304

alignment_block:

US-09-684-016-48411 x MAG39942 ..

Align seg 1/1 to: MAG39942 from: 1 to: 191

5 TATTGTAATGTGAGATGCCCTTACACCTGATGACCTAATGGCGCATG 54
|||||
115 TrrCysLysCysGluMetProTyrAsnProAspAspLeuMetValGlnCys 131
55 CGAGGGCTGCTACTGACTGG 73
|||||
131 scLuglCysLysAspTrp 137

seq_name: /SID82/gcdata/geneseq/geneseq/AA2000.DAT:MAG31123

seq_documentation_block:

ID MAG31123 standard; Protein: 196 AA.

AC MAG31123;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 37324.

KW Protein identification; signal transduction pathway; metabolic pathway;

KM hybridisation assay; genetic mapping; gene expression control; promoter;

XX Arabidopsis thaliana.

OS EPI033405-A2.

PN 06-SEP-2000.

PD 25-FEB-2000; 2000EP-0301439.

PE 25-FEB-1999; 99US-0121825.

PF 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140625.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 23-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
Quality: 138.00 Length: 23
Ratio: 6.273 Gaps: 0
Percent Similarity: 95.652 Percent Identity: 91.304

alignment block:
US-09-684-016-48411 x AAG31123 ..

Align seg 1/1 to: AAG31123 from: 1 to: 196

5 TATGTAAATGTCGATGCCTTACACCGATGACCTAATGCGCAATG 54
|||||
110 TyrcyslscyscysglumepProtyrasnProaspaspneuvelaIgincy 126
55 CGAGGCTGCACCTGACTGG 73
|||||
126 sGLuGLyCysLysasPTrp 132

seq_name: /SID2/gcgdata/geneseq/geneseq/AA2000.DAT:AAG31122

seq_documentation_block:
ID AAG31122 standard; Protein; 201 AA.
XX
AC AAG31122;
XX
DT 17-OCT-2000 (first entry)
DE
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37323.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
PD 06-SEP-2000.
XX
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138867.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139889.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150586.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160778.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.

PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
Quality: 138.00 Length: 23
Ratio: 6.273 Gaps: 0
Percent Similarity: 95.652 Percent Identity: 91.304

alignment_block:
US-09-684-016-48411 x AAG31122 ..

Align seg 1/1 to: AAG31122 from: 1 to: 201

5 TATTGTAATGTGAGATGCCTTACACCCGATGACCTAATGGCGCAATG 54
|||||
115 TyrcyslyscysglumetProtyrAsnProAspspleumetValGlnCy 131
55 CGAGGGCGCAGTGCATGCTGG 73
|||||
131 sGIuGIcysLysAspTrp 137

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG36117

seq_documentation_block:
ID AAG36117 standard; Protein; 216 AA.
XX
AC AAG36117;
XX
DT 18-OCT-2000 (first entry)
XX
DE Zea mays protein fragment SEQ ID NO: 44216.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence; corn.
XX
XX Zea mays subsp. mays.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.

```

PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147028.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149732.
PR 20-AUG-1999; 99US-0149733.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159339.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159654.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160960.
PR 22-OCT-1999; 99US-0160981.

```

```

PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
    Quality: 138.00
    Ratio: 6.273
    Percent Similarity: 95.652
    Percent Identity: 91.304

alignment_block:
US-09-684-016-48411 x AAG36117 ..
Align seg 1/1 to: AAG36117 from: 1 to: 216

5 TATTGTAATGTGAGATGCTTACACCCCTGATGACCTAATGCCCAATG 54
|||||
140 TyrcylsCysGluMetProTyrasnProAspPleuMetValGlnCy 156
55 CGAGGCTGCACCTGACTGG 73
|||||
156 sgluGlyCyslysAspTrp 162

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT: AAG05460
seq_documentation_block:
ID AAG05460 standard; Protein: 224 AA.
XX
AC AAG05460;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 1874.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.

```

PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0140991.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156556.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0158294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.

PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161820.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores: Length: 23
 Quality: 138.00
 Ratio: 6.273 Gaps: 0
Percent Similarity: 95.652 Percent Identity: 91.304

alignment_block:

US-09-684-016-48411 x AAG05460 ..

Align seg 1/1 to: AAG05460 from: 1 to: 224

5 TATTGTAATGTCAGATGCTTACACCCGTGATGACCTAATGCCCATG 54
|||||
148 TyrCysLysCysGluMetProTyrAspProAspLeuMetValGlnCys 164
55 CGAGGCGCTGCACCTGACTGG 73
|||||
164 sGIUGLYCysLysAspTrp 170

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT: AAG36116

seq_documentation_block:

ID AAG36116 standard; Protein; 338 AA.

XX AAG36116;

XX 18-OCT-2000 (first entry)

XX Zea mays protein fragment SEQ ID NO: 44215.

KM Protein identification: signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence; corn.

XX Zea mays subsp. mays.

PN EP1033405-A2.

XX 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 22-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.

```

PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0145386.
PR 02-AUG-1999; 99US-0145388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154033.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156566.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.

```

```

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
    Quality: 138.00      Length: 23
    Ratio: 6.273        Gaps: 0
Percent Similarity: 95.652  Percent Identity: 91.304

alignment_block:
US-09-684-016-48411 x AAG36116 ..

Align seg 1/1 to: AAG36116 from: 1 to: 338

5 TATGTAAATGTCGATGCCTTCAACCGATGACCTATATGCGCAAG 54
|||||
262 TyrcyslscyscysglumetProtyrAsnProaspSplumetValGncy 278
55 CGAGGCTGCACCTGACTGG 73
|||||
278 sGluGlyCysLysAspTrp 284

seq_name: /STDs2/gcgcdata/geneseq/geneseqp/AA2000.DAT:AAG04540
seq_documentation_block:
ID AAG04540 standard; Protein; 198 AA.
XX
AC AAG04540;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 604.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0123788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.

```

PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-01332484.
PR 05-MAY-1999; 99US-01332485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148584.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.

PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores:
Quality: 124.00 Length: 23
Ratio: 5.636 Gaps: 0
Percent Similarity: 95.652 Percent Identity: 78.261

alignment_block:
US-09-684-016-48411 x AAG04540 ..

Align seg 1/1 to: AAG04540 from: 1 to: 198

5 TATGTAAATGTGATGCCTTACACCCGTGATGACCTAATGGCGCAATG 54
:::|||||
111 PheCysLysCysGluMetProTyrAsnProAspAspMetValGlnCys 127
55 CGAGGCGCTGCAGTACTGC 73
|||||
127 sgluglucysSerglutr 133

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT: AAG38175

seq_documentation_block:

ID AAG38175 standard; Protein; 198 AA.

XX AAG38175;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 47057.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126284.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148345.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150966.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158363.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.

PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores: Quality: 124.00 Length: 23
 Ratio: 5.636 Gaps: 0
Percent Similarity: 95.652 Percent Identity: 78.261

alignment_block:
US-09-684-016-48411 x AAG38175 ..

Align seg 1/1 to: AAG38175 from: 1 to: 198

5 TATTGTAATGTGAGATGCCCTTACACCCGTGATGACCTAATGGCGCAATG 54
 :|||||||||||||||||||||||||||||||||||||||||
111 PhcyslyscysglumetProTyrasnProaspaspLeuMetValGlnCy 127
55 CGAGGCGTCGACACTGCTGG 73
 ||||| |||:|||||
127 sglGlucyserserGluTrp 133

seq_name: /SID2/gcgdata/geneseq/geneseqp/AA2000.DAT: AAG04539

seq_documentation_block:

ID AAG04539 standard; Protein; 228 AA.

XX AAG04539;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 603.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

PD 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

PF 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131448.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138099.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148311.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.

PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores: Quality: 124.00 Length: 23
 Ratio: 5.636 Gaps: 0
Percent Similarity: 95.652 Percent Identity: 78.261

alignment_block:

US-09-684-016-48411 x AAG04539 ..

Align seg 1/1 to: AAG04539 from: 1 to: 228

5 TATGTAAATGTAGATGCTTACAAACCGTGTGACGTAATGGCGCAATG 54
 :::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
141 PheCysIysCysGluMetProtyrAsnProAspAspLeuMetValGlnCys 157
55 CGAGGGCTGCACCTGACTGG 73
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
157 sgluGluCysserGluTrp 163

seq_name: /SITS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAG38174

seq_documentation_block:

ID AAG38174 standard; Protein: 228 AA.

AC AAG38174;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 47056.
XX
KW Protein identification: signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 26-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140635.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.

```

PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 23-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151067.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

```

```

PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

alignment_scores:      Quality: 124.00      Length: 23
                        Ratio: 5.636      Gaps: 0
Percent Similarity: 95.652      Percent Identity: 78.261

```

```

alignment_block:
US-09-684-016-48411 x MAG38174 ..

```

```

Align seq 1/1 to: MAG38174 from: 1 to: 228

```

```

5 TATGTGAATGTGATGCTTACAACTGATGACTAATGCGCATG 54
:::|||||
141 PheCysLysCysGluMetProTyrAsnProAspSpluMetValIndcy 157
55 CGAGGCTGCACGTGACTG 73
|||||
157 sGluGluCysSerCultrp 163

```

```

seq_name: /SIDS2/gcgcdata/geneseq/geneseq/AA2000.DAT: MAG40195

```

```

seq_documentation_block:

```

```

ID MAG40195 standard; Protein; 270 AA.

```

```

AC MAG40195;

```

```

DT 18-OCT-2000 (first entry)

```

```

DE Arabidopsis thaliana protein fragment SEQ ID NO: 49840.

```

```

KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

```

```

OS Arabidopsis thaliana.

```

```

PN EPI033405-A2.

```

```

PD 06-SEP-2000.

```

```

PF 25-FEB-2000; 2000EP-0301439.

```

```

XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.

```

PR 09-MAR-1999;	PR 9905-0123548.	PR 15-JUL-1999;	PR 9905-0144005.
PR 23-MAR-1999;	9905-0125788.	PR 16-JUL-1999;	9905-0144085.
PR 23-MAR-1999;	9905-0126264.	PR 16-JUL-1999;	9905-0144086.
PR 29-MAR-1999;	9905-0126765.	PR 19-JUL-1999;	9905-0144325.
PR 01-APR-1999;	9905-0127462.	PR 19-JUL-1999;	9905-0144331.
PR 06-APR-1999;	9905-0128234.	PR 19-JUL-1999;	9905-0144332.
PR 08-APR-1999;	9905-0128715.	PR 19-JUL-1999;	9905-0144333.
PR 16-APR-1999;	9905-0129845.	PR 19-JUL-1999;	9905-0144334.
PR 19-APR-1999;	9905-0130077.	PR 19-JUL-1999;	9905-0144335.
PR 21-APR-1999;	9905-0130449.	PR 20-JUL-1999;	9905-0144332.
PR 23-APR-1999;	9905-0130510.	PR 20-JUL-1999;	9905-0144632.
PR 28-APR-1999;	9905-0130891.	PR 21-JUL-1999;	9905-0144884.
PR 30-APR-1999;	9905-0132048.	PR 21-JUL-1999;	9905-0145086.
PR 03-MAY-1999;	9905-0132407.	PR 21-JUL-1999;	9905-0145088.
PR 04-MAY-1999;	9905-0132484.	PR 22-JUL-1999;	9905-0145087.
PR 05-MAY-1999;	9905-0132485.	PR 22-JUL-1999;	9905-0145087.
PR 06-MAY-1999;	9905-0132486.	PR 22-JUL-1999;	9905-0145089.
PR 06-MAY-1999;	9905-0132487.	PR 22-JUL-1999;	9905-0145192.
PR 07-MAY-1999;	9905-0132863.	PR 23-JUL-1999;	9905-0145145.
PR 11-MAY-1999;	9905-0134256.	PR 23-JUL-1999;	9905-0145218.
PR 14-MAY-1999;	9905-0134221.	PR 23-JUL-1999;	9905-0145224.
PR 14-MAY-1999;	9905-0134219.	PR 26-JUL-1999;	9905-0145276.
PR 14-MAY-1999;	9905-0134370.	PR 27-JUL-1999;	9905-0145913.
PR 18-MAY-1999;	9905-0134768.	PR 27-JUL-1999;	9905-0145918.
PR 19-MAY-1999;	9905-0134941.	PR 28-JUL-1999;	9905-0145919.
PR 20-MAY-1999;	9905-0135124.	PR 28-JUL-1999;	9905-0145951.
PR 21-MAY-1999;	9905-0135352.	PR 02-AUG-1999;	9905-0146386.
PR 24-MAY-1999;	9905-0135629.	PR 02-AUG-1999;	9905-0146388.
PR 25-MAY-1999;	9905-0136021.	PR 02-AUG-1999;	9905-0146389.
PR 27-MAY-1999;	9905-0136392.	PR 03-AUG-1999;	9905-0147038.
PR 28-MAY-1999;	9905-0136782.	PR 04-AUG-1999;	9905-0147204.
PR 01-JUN-1999;	9905-0137222.	PR 05-AUG-1999;	9905-0147302.
PR 03-JUN-1999;	9905-0137528.	PR 05-AUG-1999;	9905-0147260.
PR 04-JUN-1999;	9905-0137502.	PR 06-AUG-1999;	9905-0147303.
PR 07-JUN-1999;	9905-0137724.	PR 06-AUG-1999;	9905-0147416.
PR 08-JUN-1999;	9905-0138094.	PR 09-AUG-1999;	9905-0147493.
PR 10-JUN-1999;	9905-0138540.	PR 09-AUG-1999;	9905-0147935.
PR 10-JUN-1999;	9905-0138847.	PR 11-AUG-1999;	9905-0148171.
PR 14-JUN-1999;	9905-0139119.	PR 11-AUG-1999;	9905-0148319.
PR 16-JUN-1999;	9905-0139452.	PR 12-AUG-1999;	9905-0148341.
PR 16-JUN-1999;	9905-0139453.	PR 13-AUG-1999;	9905-0148565.
PR 17-JUN-1999;	9905-0139452.	PR 13-AUG-1999;	9905-0148684.
PR 18-JUN-1999;	9905-0139454.	PR 16-AUG-1999;	9905-0149328.
PR 18-JUN-1999;	9905-0139455.	PR 17-AUG-1999;	9905-0149175.
PR 18-JUN-1999;	9905-0139456.	PR 17-AUG-1999;	9905-0149175.
PR 18-JUN-1999;	9905-0139457.	PR 18-AUG-1999;	9905-0149426.
PR 18-JUN-1999;	9905-0139458.	PR 20-AUG-1999;	9905-0149722.
PR 18-JUN-1999;	9905-0139460.	PR 20-AUG-1999;	9905-0149723.
PR 18-JUN-1999;	9905-0139460.	PR 23-AUG-1999;	9905-0149902.
PR 18-JUN-1999;	9905-0139461.	PR 23-AUG-1999;	9905-0149902.
PR 18-JUN-1999;	9905-0139462.	PR 25-AUG-1999;	9905-0150566.
PR 18-JUN-1999;	9905-0139463.	PR 25-AUG-1999;	9905-0150566.
PR 18-JUN-1999;	9905-0139750.	PR 26-AUG-1999;	9905-0150884.
PR 18-JUN-1999;	9905-0139763.	PR 27-AUG-1999;	9905-0151065.
PR 21-JUN-1999;	9905-0139817.	PR 27-AUG-1999;	9905-0151066.
PR 22-JUN-1999;	9905-0139899.	PR 30-AUG-1999;	9905-0151303.
PR 23-JUN-1999;	9905-0140353.	PR 30-AUG-1999;	9905-0151303.
PR 23-JUN-1999;	9905-0140354.	PR 31-AUG-1999;	9905-0151438.
PR 24-JUN-1999;	9		

PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores: Quality: 124.00 Length: 23
 Ratio: 5.636 Gaps: 0
Percent Similarity: 95.652 Percent Identity: 78.261

alignment_block:

US-09-684-016-48411 x AAG40195 ..

Align seg 1/1 to: AAG40195 from: 1 to: 270

5 TATTGTAATGTGATGCGCTTACACCCGTGACCTAATGGCGCATG 54
 :::|||||
183 PheCysLysCysGluMetProTyrAsnProAspAspLeuMetValGlnCys 199
55 CGAGGCGTCGACTGACTGG 73
 |||| |
199 sgluGlnCysSerGluTrrp 205

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA2000.DAT: AAG40194

seq_documentation_block:

ID AAG40194 standard; Protein: 357 AA.

XX AAG40194;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 49839.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-01332407.
PR 05-MAY-1999; 99US-0133484.
PR 06-MAY-1999; 99US-0133485.
PR 06-MAY-1999; 99US-0133486.
PR 07-MAY-1999; 99US-01332863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0138119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.

PD 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148341.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148688.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151303.
PR 30-AUG-1999; 99US-0151308.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.

XX AAB42371;
 AC
 XX 08-FEB-2001 (first entry)
 DT
 XX
 DE Human ORFX ORF2135 polypeptide sequence SEQ ID NO:4270.
 XX
 KW Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
 KW vulnereary; antiparietal; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antirheumatic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;
 KW antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
 KW antineoplastic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antineoplastic disease; coagulation;
 KW thrombosis; contraceptive.
 KW
 XX Homo sapiens.
 OS
 PN MO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000MO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shimkets RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 DR N-PSDB; AAC76580.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 11: Page 3476-3478; 5507pp; English.
 XX
 CC AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytosolic; hepatotropic; vulnereary;
 CC antiparietal; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antirheumatic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antineoplastic; antibacterial; antiviral; antifungal; antineoplastic;
 CC antihypertensive; and antineoplastic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, diabetes mellitus,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antineoplastic disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 XX Sequence 1084 AA;

alignment_scores: 74.00 Length: 23
 quality: 4.625 Gaps: 0
 Percent Similarity: 69.565 Percent Identity: 43.478
 alignment_block:
 US-09-684-016-48411 x AAB42371 ..
 Align seq 1/1 to: AAB42371 from: 1 to: 1084
 5 TATGTAAAGTGGATGCGCTTACACCGCTGAGGACCTAATGCGCAATG 54
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 67 Tyrosylleucylsargineprotyrvaltharphenylleucyl 83
 55 CGAGGCTGCACTGACTG 73
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 83 saspmeccysglinsprr 89
 seq_name: /STD2/gcgdata/geneseq/geneseq/AA2001.DAT:AA40982
 seq_documentation_block:
 ID AAM40982 standard; Protein: 384 AA.
 AC
 XX AAM40982;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 5913.
 XX
 KW Human; nootropic; immunosuppressant; cytosolic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 KW
 XX Homo sapiens.
 OS
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000MO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR WPI: 2001-442253/47.
 DR N-PSDB; AAI60138.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Example 2: SEQ ID NO 5913; 10078pp; English.
 XX
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AA42213) with nootropic,
 CC immunosuppressant and cytosolic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and

CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX Sequence 384 AA;

alignment_scores:

Quality: 66.50 Length: 29
 Ratio: 4.433 Gaps: 2
 Percent Similarity: 51.724 Percent Identity: 48.276

alignment_block:

US-09-684-016-48411 x AAM40982 ..

Align seg 1/1 to: AAM40982 from: 1 to: 384

5 TATTGTAATGTGAGATGCTTACACCTGAT..... 37
 ||||| |||: ||||| |||||
 137 TyrCysIleCysLysArgProTyr...ProAspProGluAspGluIlePr 152
 38 .GACCTAATGGCGCAATGGAGGGCTGCACCTGACTGG 73
 ||| ||| ||||| ||| |||||
 152 oAspGluMetIleGlnCysValValCysGluAspTrp 164

seq_name: /SIDS2/gcgdata/geneseq/geneseq/Aa2001.DAT:AAM39196

seq_documentation_block:

ID AAM39196 standard; Protein: 425 AA.

XX AAM39196;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 2341.

KW Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

XX Homo sapiens.

XX WO200153312-A1.

PD 26-JUL-2001.

PE 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.

XX (HXSE-) HXSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang Z, Wang J, Wehrman T, Xu C, Xue AU, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX

DR WPI: 2001-442253/47.
 DR N-PSDB: AAI58352.

PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -

XX Example 4; SEQ ID NO 2341; 10078bp; English.

CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with noctropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX Sequence 425 AA;

alignment_scores:

Quality: 66.50 Length: 29
 Ratio: 4.433 Gaps: 2
 Percent Similarity: 51.724 Percent Identity: 48.276

alignment_block:

US-09-684-016-48411 x AAM39196 ..

Align seg 1/1 to: AAM39196 from: 1 to: 425

5 TATTGTAATGTGAGATGCTTACACCTGAT..... 37
 ||||| |||: ||||| |||||
 134 TyrCysIleCysLysArgProTyr...ProAspProGluAspGluIlePr 149
 38 .GACCTAATGGCGCAATGGAGGGCTGCACCTGACTGG 73
 ||| ||| ||||| ||| |||||
 149 oAspGluMetIleGlnCysValValCysGluAspTrp 161

seq_name: /SIDS2/gcgdata/geneseq/geneseq/Aa2001.DAT:AAB60498

seq_documentation_block:

ID AAB60498 standard; Protein: 425 AA.

XX AAB60498;

DT 24-APR-2001 (first entry)

DE Human cell cycle and proliferation protein CCYPR-46, SEQ ID NO:46.

KW Cell cycle and proliferation protein; CCYPR; human; agonist;
 KW antagonist; gene therapy; detection; gene therapy;
 KW transgenic animal disease model; immune disorder;
 KW developmental disorder; cell signalling disorder;
 KW cell proliferative disorder; cancer; tumour; anaemia; epilepsy;
 KW arteriosclerosis; asthma; allergy; diabetes mellitus;
 KW menstrual cycle disorder; bacterial infection.

XX Homo sapiens.

XX WO200107471-A2.

XX 01-FEB-2001.

XX 21-JUL-2000; 2000WO-US19948.


```
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
```

```
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.
```

```
alignment_scores:
Quality: 65.00 Length: 14
Ratio: 5.417 Gaps: 0
Percent Similarity: 85.714 Percent Identity: 78.571
```

```
alignment_block:
US-09-684-016-48411 x AAG47973 ..
```

```
Align seg 1/1 to: AAG47973 from: 1 to: 193
```

```
5 TATTGTAATGTCGATGCTTACAACTGATGACCTAATG 46
|||||
110 TYRTRYLYSCYSGLUMETPROTYRASNRSERASPGIUMET 123
```

```
seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT:AAV04323
```

```
seq_documentation_block:
```

```
ID AAV04323 standard; Protein: 238 AA.
```

```
XX AAV04323;
```

```
DT 18-JUN-1999 (first entry)
```

```
XX Human secreted protein SEO ID NO:69.
```

```
XX Human; secreted protein; cancer; tumour; developmental abnormality;
```

foetal deficiency; blood disorder; immune system disorder; inflammation;
 autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
 schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
 atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
 digestive disorder; endocrine disorder; infection; AIDS.
 XX Homo sapiens.
 OS
 XX
 PN W09910363-A1.
 PD
 PD 04-MAR-1999.
 XX
 PF 27-AUG-1998; 98WO-US17709.
 XX
 XX 29-AUG-1997; 97US-0056271.
 PR 29-AUG-1997; 97US-0056073.
 PR 29-AUG-1997; 97US-0056247.
 PR 29-AUG-1997; 97US-0056270.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Fan P, Kyaw H, Rosen CA, Ruben SM, Wei YF;
 XX
 DR WPI: 1999-190585/16.
 XX
 PT New isolated human genes and the secreted polypeptides they encode
 PT - useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX
 PS Disclosure: Page 18; 170pp: English.
 XX
 CC AAX30145 to AAX30173 represent 29 isolated human secreted protein genes.
 CC AAX04293 to AAY04321 represent the secreted proteins encoded by the 29
 CC human genes. The genes and their corresponding secreted polypeptides are
 CC useful for preventing, treating or ameliorating medical conditions,
 CC e.g. by protein or gene therapy. Also pathological conditions can be
 CC diagnosed by determining the amount of the new polypeptides in a sample
 CC or by determining the presence of mutations in the new genes. Specific
 CC uses are described for each of the 29 genes, based on which tissues they
 CC are most highly expressed in, and include developing products for the
 CC diagnosis or treatment of cancer, tumours, developmental abnormalities
 CC and foetal deficiencies, blood disorders, diseases of the immune system,
 CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive
 CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
 CC disorders, digestive/endocrine disorders, cardiovascular disorders, kidney
 CC disorders, digestive/endocrine disorders, infections and AIDS. The
 CC polypeptides are also useful for identifying their binding partners.
 CC The sequences given in AAX30174 to AAX30182 and AAY04322 to AAY04334 are
 CC used in the exemplification of the present invention.
 CC
 CC Sequence 238 AA;
 CC

```

alignment_scores:
    quality: 60.00
    ratio: 4.000
    percent_similarity: 65.217
    percent_identity: 34.783

alignment_block:
    US-09-684-016-48411 x AAY04323 ..

Align seg 1/1 to: AAY04323 from: 1 to: 238

5 TATTGTAATGTGAGATGTCCTTACAACTGATGACCTTATAGGCGCAATG 54
   ||||| ||||: |||||::: ::::: ||
61 TTYCYSILIECYASARPhrProTyrAspGlnSerGlnPhenylleuYcy 77
   CGAGGCGCTGCCTACTGACTGG 73
   |::: |||
77 sAsparGcYsGlnAsntrIIP 83

seq_name: /SID52/9/cgdata/geneseq/geneseq/AAY57453

```

```
seq_documentation_block:
ID AAY57453 standard; Protein; 2781 AA.
XX
XX AAY57453;
AC
XX
DE 22-FEB-2000 (first entry)
XX
XX Human transcriptional regulatory factor SEQ ID NO:10.
XX
XX Human; transcriptional regulatory factor; TCOAL; BLAST detection;
KW bromo-domain; cell proliferation; cancer.
XX
XX Homo sapiens.
OS
XX W09597143-A1.
PN
XX 11-NOV-1999.
PD
XX
XX 30-APR-1999; 99WO-JP02340.
PF
XX
XX 30-APR-1998; 98JP-0137631.
PR
XX
XX (CHUG-) CHUGAI RES INST MOLECULAR MEDICINE INC.
PA
XX Jones MH;
PI
XX WPI: 2000-052940/04.
DR
XX N-PSDB; AAZ39033.
XX
XX Transcriptional regulatory factor containing a bromo domain and gene
PT TCOAL encoding it -
PS
XX Claim 1; Page 139-151; 154pp; Japanese.
XX
CC The present sequence represents a human transcriptional regulatory factor
CC containing a bromo domain. The factor interacts with proteins involved
CC in the chromatin-mediated transcription regulation mechanism. It binds
CC to hSNF2H, hSNF2L and MCOA-62/Skip. It can be used for screening
CC compounds binding to it and acting as agonists or antagonists, which
CC are potentially useful for the treatment and prevention of cancer and
CC other cell proliferation disorders.
XX
XX Sequence 2781 AA;
S0
alignment_scores:
Quality: 60.00 Length: 23
Ratio: 4.286 Gaps: 0
Percent Similarity: 60.870 Percent Identity: 34.783
alignment_block:
US-09-684-016-48411 x AAY57453 ..
Align seg 1/1 to: AAY57453 from: 1 to: 2781
5 TATGTAAATGTTGAGATGCCCTTACACCCTGATGACTAATGGCGCAATG 54
||||| ||||| |||||::: :::: ||
2604 TyrCysIleCysLysThrProTyrAspGluSerLysPheTyrIleGlyCys 2620
55 CGAGGGCTGCACCTGACTGG 73
1::: ||| :::|||
2620 SASPARGYCSGLNASNTRP 2626
seq_name: /SIDIS2/gcgdata/geneseq/geneseqp/AA2000.DAT:AAY57452
seq_documentation_block:
ID AAY57452 standard; Protein; 2907 AA.
XX
XX AAY57452;
AC
XX
XX 22-FEB-2000 (first entry)
XX
XX Human transcriptional regulatory factor SEQ ID NO:1.
```

XX	06-SEP-2000.	2000EP-0301439
PD		
XX		
PF	25-FEB-2000;	2000EP-0301439
XX		
PR	25-FEB-1999;	9905-01231825
PR	03-MAR-1999;	9905-0123180
PR	05-MAR-1999;	9905-01233548
PR	25-MAR-1999;	9905-0125789
PR	25-MAR-1999;	9905-01256564
PR	29-MAR-1999;	9905-0126785
PR	01-APR-1999;	9905-0127462
PR	01-APR-1999;	9905-01282534
PR	08-APR-1999;	9905-01287114
PR	16-APR-1999;	9905-01298415
PR	19-APR-1999;	9905-0130077
PR	21-APR-1999;	9905-01300449
PR	22-APR-1999;	9905-0130510
PR	23-APR-1999;	9905-0130891
PR	28-APR-1999;	9905-0131449
PR	30-APR-1999;	9905-01320408
PR	30-APR-1999;	9905-0132407
PR	04-MAY-1999;	9905-01324884
PR	05-MAY-1999;	9905-01332885
PR	06-MAY-1999;	9905-01332486
PR	07-MAY-1999;	9905-0132487
PR	11-MAY-1999;	9905-01324556
PR	14-MAY-1999;	9905-01334218
PR	14-MAY-1999;	9905-01334219
PR	14-MAY-1999;	9905-01334220
PR	14-MAY-1999;	9905-01334271
PR	18-MAY-1999;	9905-01347681
PR	19-MAY-1999;	9905-01347461
PR	20-MAY-1999;	9905-0135124
PR	21-MAY-1999;	9905-01353553
PR	24-MAY-1999;	9905-01355629
PR	25-MAY-1999;	9905-0136021
PR	27-MAY-1999;	9905-01367822
PR	28-MAY-1999;	9905-01372722
PR	01-JUN-1999;	9905-0137528
PR	04-JUN-1999;	9905-01377502
PR	07-JUN-1999;	9905-01377244
PR	08-JUN-1999;	9905-01380940
PR	10-JUN-1999;	9905-0138540
PR	10-JUN-1999;	9905-0138847
PR	14-JUN-1999;	9905-0139119
PR	16-JUN-1999;	9905-01394512
PR	16-JUN-1999;	9905-01394653
PR	17-JUN-1999;	9905-01394652
PR	18-JUN-1999;	9905-0139454
PR	18-JUN-1999;	9905-0139455
PR	18-JUN-1999;	9905-0139456
PR	18-JUN-1999;	9905-0139457
PR	18-JUN-1999;	9905-0139458
PR	18-JUN-1999;	9905-0139459
PR	18-JUN-1999;	9905-0139750
PR	18-JUN-1999;	9905-0139763
PR	21-JUN-1999;	9905-0139617
PR	22-JUN-1999;	9905-0139699
PR	23-JUN-1999;	9905-0140355
PR	24-JUN-1999;	9905-0140354
PR	24-JUN-1999;	9905-0140695
PR	28-JUN-1999;	9905-0140823
PR	29-JUN-1999;	9905-0140981
PR	30-JUN-1999;	9905-0141287
PR	01-JUL-1999;	9905-0141542
PR	01-JUL-1999;	9905-0141543

PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145917.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.

PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157173.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores: Quality: 57.00 Length: 24
 Ratio: 3.167 Gaps: 0
Percent Similarity: 75.000 Percent Identity: 33.333

alignment_block:

US-09-684-016-48411/rev x AAG51002 ..

Align seg 1/1 to: AAG51002 from: 1 to: 480

104 AGAGTAAGTATTCACCTTACATGTGAACCTTACAGTCAGTCAGCCCTCG 55
||||:||||: ::||| ||| ::|||
161 LysleuAhsnleuGlutIleHisIleTyValThrArgGlnSerGluProCys 177
54 CATTGGCCCATTTAGCTCATCAG 33
:::||||| |||||:::
177 sleuGlyProTyGlyHisGlu 184

seq_name: /SIDS2/gcdata/geneseq/geneseq/AA2000.DAT:AAG51001

seq_documentation_block:

ID AAG51001 standard; Protein: 567 AA.

AC AAG51001;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 64687.

XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137232.
PR 03-JUN-1999; 99US-0137502.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147319.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.

PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149358.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.

PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

alignment_scores: 57.00 Length: 24
Quality: 3.167 Gaps: 0
Ratio: 3.167
Percent Similarity: 75.000 Percent Identity: 33.333

alignment_block:

US-09-684-016-48411/rev x AAG51000 ..

Align seg 1/1 to: AAG51000 from: 1 to: 614

104 AAGATAGATATACCTTACATGTGAACCTTACACTGACGACCCCTCG 55
||||:||||: |||
295 lYsleuAsnleuGlnIleHisIleTyValTrArgInserGlnProcy 311
:||||: |||

54 CATTCGCCCATTAGGTATCAG 33
:||||: |||

311 slenclYProTyRGLYHISGLN 318

seq_name: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT: AAB43724

seq_documentation_block:

ID AAB43724 standard; Protein: 181 AA.

XX AAB43724;

XX 08-FEB-2001 (first entry)

DT XX

DE Human cancer associated protein sequence SEQ ID NO:1169.

KW Human; cancer associated gene; cancer antigen; detection; cancer;
KM diagnosis; cytostatic; proliferative; vulnerability; immunomodulator;
KM antidiabetic; antitachymatic; antirheumatic; antiarthritic; antiviral;
KM antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KM dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KM vasotropic; antiporiatic; antiangiogenic; gene therapy; inflammation;
KM immune disorder; haematopoietic cell disorder; autoimmune disorder;
KM allergic reaction; graft versus host disease; organ rejection;
KM haemostatic; thrombolytic; cardiovascular disorder; infection;
KW neurological disease; drug screening.

OS Homo sapiens.
XX
XX WO200055350-A1.
PN
XX
XX 21-SEP-2000.
PD
XX
XX 08-MAR-2000; 2000WO-US05882.
PF
XX
XX 12-MAR-1999; 99US-0124270.
PR
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX Rosen CA, Ruben SM;
PI
XX
XX WPI; 2000-587533/55.
DR
XX
XX N-PSDB; AAC7933.
PT
XX
XX Novel isolated nucleic acids comprising sequences encoding peptides
PT useful for treating or diagnosing e.g. cancer -

Claim 11; Page 1790-1791; 2352pp; English.

CC AAC7607 to AAC7844 encode the human cancer associated proteins given
CC in AAB3398 to AAB4439. The proteins can have activities based on the
CC tissues and cells the genes are expressed in. Example of activities
CC include: cytostatic; proliferative; vulnerability; immunomodulator;
CC antidiabetic; antitachymatic; antirheumatic; antiarthritic;
CC antiinflammatory; antithyroid; antiallergic; antibacterial; antiviral;
CC dermatological; neuroprotective; cardiant; thrombolytic; coagulant;
CC nootropic; vasotropic; antiporiatic and antiangiogenic. The
CC polynucleotides and polypeptides can be used for preventing, treating or
CC ameliorating medical conditions and diagnosing pathological conditions.
CC Polynucleotides, polypeptides, antibodies, agonists and antagonists from
CC the present invention may be used to treat immune disorders by activating
CC or inhibiting the proliferation, differentiation or mobilisation of
CC immune cells, to treat disorders of haematopoietic cells, autoimmune
CC disorders, allergic reactions, graft versus host disease and organ
CC rejection, modulate haemostatic or thrombolytic activity, modulate
CC inflammation, cancers, cardiovascular disorders, neurological disease and
CC bacterial or viral infections. The peptides, nucleotides, antibodies
CC agonists and antagonists may be also be used in drug screens. AAC7849 to
CC AAC7847 and AAB44240 represent sequences used in the exemplification of
CC the present invention.
XX
XX

SQ Sequence 181 AA:

Alignment-scores:
Quality: 56.50 Length: 23
Ratio: 3.531 Gaps: 1
Percent Similarity: 69.565 Percent Identity: 39.130

Alignment_block:
US-09-684-016-48411 x AAB43724 ..

Align seg 1/1 to: AAB43724 from: 1 to: 181

5 TATTGTAATGTGAGATGCTTACACCCGTATACCTAATGGCGCATG 54
||||| |||:|||||:||||| :| :||| ||
26 TTYCysIIleCYsArGInPrOHISAsn...AsnArGPrHeMeIIleCYsCY 41

```

seq_name: /SID52/gcdata/geneseq/geneseqp/AA2000.DAT:AA67579
seq_documentation_block:
ID      AA67579 standard; Protein: 562 AA.
XX
AC      AAY67579;
XX
DT      19-JUN-2000 (first entry)
XX
DE      Human death inducer-oblitertor 1 (DIO-1) polypeptide.
XX
DE      Death inducer-oblitertor 1; DIO-1; cell death; cancer; tumour; human;
KM      autoimmune disease; cytostatic; immunosuppressive; antidiabetic;
KW      antineumatic; antiinflammatory; antiproliferative.
XX
OS      Homo saplens.
XX
FH      Key Location/Oualifiers
FT      Region 165..172
FT      /note= "NLS sequence"
FT      Region 185..193
FT      /note= "NLS sequence"
FT      Region 271..288
FT      /note= "zinc finger motif"
FT      Region 293..320
FT      /note= "zinc finger motif"
XX
PN      WO200015787-A1.
PD
PD      23-MAR-2000.
XX
XX      PF 10-SEP-1999; 99WO-GBO3019.
XX
PR      10-SEP-1998; 98SE-0003069.
PR      17-SEP-1998; 98US-0100873.
XX
XX      (CNSJ ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.
PA      (BANN/) BANNERMAN D G.
XX
PI      Alonso CM, Domingo DG, Grandien A, Leonardo E, Martinez P;
DR      WPI: 2000-271426/23.
DR      N-PSDB; AA290578.
XX
PT      New DNA encoding human and murine death inducer-oblitertor 1
PT      polypeptides, useful in the treatment of cancer, autoimmune diseases,
PT      diabetes, rheumatoid arthritis, benign tumors, malignant tumors and
PT      hyperproliferative skin disorders -
XX
SS      Claim 6; Fig 1C; 27pp; English.
XX
CC      The invention provides nucleic acids encoding the human and murine death
CC      inducer-oblitertor 1 (DIO-1) polypeptides. The polypeptides can be
CC      expressed by standard recombinant methodology. The DIO-1 polypeptides,
CC      agonists and antagonists are used as a medicament for treating diseases
CC      characterized by an alteration in cell death or by hyperproliferation,
CC      e.g. cancer, autoimmune diseases, diabetes, rheumatoid arthritis, benign
CC      tumours, malignant tumours or hyperproliferative skin disorders. They are
CC      also useful in the treatment of metabolic, proliferative or inflammatory
CC      conditions. The present sequence represents the human DIO-1 polypeptide.
XX
SQ      Sequence 562 AA;

Alignment_scores:
Quality: 56.50 Length: 23
Ratio: 3.531 Gaps: 1
Percent Similarity: 69.565 Percent Identity: 39.130

```

alignment_block:
US-09-684-016-48411 x AAY67579 ..

Align seg 1/1 to: AAY67579 from: 1 to: 562

```

5 TATGTAAATGTGAGATGCTTACACCCGTGATGACCTAATGCGCAATG 54
||||| |||:|||||:||||| |||: |||
270 TyrcystleCySargGlnProHisAsn...AsnArgPheMetIleCyScy 285
55 CGAGGGCTGCACACTGACTGG 73
|::| ||| :|||
285 sAspArgCySglucIutrp 291

```

seq_name: /SIDIS2/gcgdata/geneseq/geneseqp/AA2001.DAT.AAB93638

seq_documentation_block:

```

ID AAB93638 standard; Protein; 562 AA.
XX
AC AAB93638;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:13130.
XX
KM Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-0116126.
XX
PR 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR WPI; 2001-318749/34.
XX
PT Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
PS Claim 8; SEQ ID 13130; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesizing 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination of
XX the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesizing polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

```

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.

XX Sequence 562 AA;

alignment_scores:

Quality:	Length:
Ratio:	23
Percent Similarity:	Gaps: 1
Percent Identity:	39.130

alignment_block:

US-09-684-016-48411 x AAB93638 ..

Align seg 1/1 to: AAB93638 from: 1 to: 562

```

5 TATGTAAATGTGAGATGCTTACACCCGTGATGACCTAATGCGCAATG 54
||||| |||:|||||:||||| |||: |||
270 TyrcystleCySargGlnProHisAsn...AsnArgPheMetIleCyScy 285
55 CGAGGGCTGCACACTGACTGG 73
|::| ||| :|||
285 sAspArgCySglucIutrp 291

```

seq_name: /SIDIS2/gcgdata/geneseq/geneseqp/AA2000.DAT.AAY67580

seq_documentation_block:

```

ID AAY67580 standard; Protein; 614 AA.
XX
AC AAY67580;
XX
DT 19-JUN-2000 (first entry)
XX
DE Murine death inducer-obliteror 1 (DIO-1) polypeptide.
XX
KM Death inducer-obliteror 1; DIO-1; cell death; cancer; tumour; mouse;
KM autoimmune disease; cytostatic; immunosuppressive; antidiabetic;
KM antineumatic; antiinflammatory; antiproliferative.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FH Region 162..170
FT /note="NLS sequence"
FT Region 182..190
FT /note="NLS sequence"
FT Region 262..285
FT /note="zinc finger motif"
FT Region 290..315
FT /note="zinc finger motif"
XX
XX WO200015787-A1.
XX
PD 23-MAR-2000.
XX
PF 10-SEP-1999; 99WO-GB03019.
XX
PR 10-SEP-1998; 98SE-0003069.
PR 17-SEP-1998; 98US-0100873.
XX
PA (CNSI ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.
PA (BANN ) BANNERMAN D G.
XX
PI Alonso CM, Domingo DG, Grandien A, Leonardo E, Martinez P;
DR N-PSDB; AA290579.
XX
XX New DNA encoding human and murine death inducer-obliteror 1
XX polypeptides, useful in the treatment of cancer, autoimmune diseases,
XX diabetes, rheumatoid arthritis, benign tumors, malignant tumors and

```



```
XX Hayden MR, Kalchman M;
PI
XX
DR WPI: 1997-297870/27.
DR N-PSDB; AAT67187.
XX
PT cDNA encoding Huntington's disease interacting protein - useful to
PT ameliorate effects of disease in patient expressing Huntingtin
PT protein with expanded CAG repeat region
XX
PS Claim 4; Page 28-31; 39pp: English.
XX
CC Huntingtin interacting protein HIP1 (AAW18030) is a brain-specific
CC protein that interacts differently with the gene product of a
CC normal (16 CAG repeat) and an expanded (over 44 CAG repeat)
CC Huntington's disease (HD) gene, and which facilitates incorporation
CC of huntingtin into brain cell membranes. Its amino acid sequence
CC was deduced from a human frontal cortex cDNA clone (AAT67187). The
CC effects of HD in a patient expressing huntingtin protein with an
CC expanded polyglutamine tract can be ameliorated by increasing the
CC amount of expressed HIP1 in the brain using gene therapy
CC approaches. Modified forms of HIP1 which bind more effectively to
CC expanded HD protein can be used to convert the expanded protein
CC into a functional molecule.
XX
SQ Sequence 914 AA:

alignment_scores:
  Quality: 54.00 Length: 38
  Ratio: 2.077 Gaps: 1
  Percent Similarity: 68.421 Percent Identity: 36.842

alignment_block:
  US-09-684-016-48411/rev x AAW18030 ..

Align seg 1/1 to: AAW18030 from: 1 to: 914

119 GTGTGGAATAATGAGATAGATACCTTACATGGAACCTTACCAG 70
   |||:::|||||::: |||::: |||:::
79 ValPheAsnSerLeuAspMetSerArgSerValSerValThrAlaIaIaGI 95
69 TCAGTGCAGCCCTGCATTCGCCCATTAAGTCATCAGGGTTGTAAGCAT 20
   :||| |||:::||||| |||::: |||:::
95 yGlnCys.....ArgLeuAlaProLeuIleGlnValIleIleuAspCys 110
19 CTCACATTTTACAT 6
   |||:::|||||:::
110 erHisLeuTYrAsp 114

seq_name: /SID52/gcgdata/geneseq/geneseqp/AA2000.DAT.AAY59269
seq_documentation_block:
ID AAY59269 standard; Protein; 914 AA.
XX
AC AAY59269;
XX
DT 17-APR-2000 (first entry)
XX
DE Human huntingtin-interacting protein (HIP1).
XX
KW Huntingtin Interacting Protein; HIP; death effector domain; DED; human;
KW apoptosis; HIP-apoptosis modulating protein; cell death; gene therapy;
KW Huntington's disease; neurotropic; anticonvulsant; cyostatic;
KW chromosome 7q11.23.
XX
XX Homo sapiens.
XX
XX OS
XX PN WO960986-A2.
XX
XX PD 02-DEC-1999.
XX
XX PF 27-MAY-1999; 99WO-US11743.
```

```
XX
PR 27-MAY-1998; 98US-0085199.
XX
PA (UYER-) UNIV BRITISH COLUMBIA.
PA (MERI) MERCK FROST CANADA INC.
XX
PI Kalchman M, Hayden MR, Hackam A, Chopra V, Nicholson DW;
PI Vallaincourt JP, Rasper DM;
XX
DR WPI: 2000-097055/08.
DR N-PSDB; AAZ58746.
XX
PT Novel proteins useful for treating Huntington's disease by gene therapy
PT techniques, and cancers
XX
PS Claim 9; Page 53-57; 91pp: English.
XX
CC The invention relates to Huntingtin Interacting Protein (HIP), that
CC includes a death effector domain (DED), suggesting apoptotic function.
CC Proteins with DED (referred as HIP-apoptosis modulating proteins) are
CC useful for inducing apoptotic death in cells. The HIP is a logical target
CC for therapy in Huntington's disease since it has a differential apoptotic
CC activity, modulated by interaction with Huntingtin having normal and
CC expanded repeats. HIP is also used as a therapeutic agent to introduce
CC apoptosis in cancer cells. Increasing expression of normal (non-expanded)
CC Huntingtin or the HIP-apoptotic modulator-binding portion, a modified
CC HIP-apoptotic modulator in which the DED has been deleted, is helpful for
CC treating Huntington's disease. The present sequence represents the human
CC HIP1.
XX
SQ Sequence 914 AA:

alignment_scores:
  Quality: 54.00 Length: 38
  Ratio: 2.077 Gaps: 1
  Percent Similarity: 68.421 Percent Identity: 36.842

alignment_block:
  US-09-684-016-48411/rev x AAY59269 ..

Align seg 1/1 to: AAY59269 from: 1 to: 914

119 GTGTGGAATAATGAGATAGATACCTTACATGGAACCTTACCAG 70
   |||:::|||||::: |||::: |||:::
79 ValPheAsnSerLeuAspMetSerArgSerValSerValThrAlaIaIaGI 95
69 TCAGTGCAGCCCTGCATTCGCCCATTAAGTCATCAGGGTTGTAAGCAT 20
   :||| |||:::||||| |||::: |||:::
95 yGlnCys.....ArgLeuAlaProLeuIleGlnValIleIleuAspCys 110
19 CTCACATTTTACAT 6
   |||:::|||||:::
110 erHisLeuTYrAsp 114

seq_name: /SID52/gcgdata/geneseq/geneseqp/AA2000.DAT.AAY59270
seq_documentation_block:
ID AAY59270 standard; Protein; 1090 AA.
XX
AC AAY59270;
XX
DT 17-APR-2000 (first entry)
XX
DE Human huntingtin-interacting protein (HIP1).
XX
KW Huntingtin Interacting Protein; HIP; death effector domain; DED; human;
KW apoptosis; HIP-apoptosis modulating protein; cell death; gene therapy;
KW Huntington's disease; neurotropic; anticonvulsant; cyostatic;
KW chromosome 7q11.23.
XX
XX Homo sapiens.
XX
XX OS
```


CC secreted protein of the invention.
XX
SQ Sequence 132 AA;

alignment_scores:
Quality: 53.50 Length: 51
Ratio: 2.058 Gaps: 4
Percent Similarity: 50.980 Percent Identity: 35.294

alignment_block:

US-09-684-016-48411/rev x AAE03341 ..

Align seg 1/1 to: AAE03341 from: 1 to: 132

```
116 TGGATTAATTCAGATAGTATCACCCTTACATGTGAAA...CTTACCAG 70
|||||  |||  |||:  |||:  |||  |||
50 TTPASNIleuLeuTyrSerValGlySerIysValSerGlyLeuCysSe 66
69 TCAGTCAGCAGC.....CCTCGCA 53
||:|||||
66 TAsncysSerLeuValProGlyValAlaIAlaHisThrCysAsnProLysV 83
52 TTGGCCCATTAAGTCATCAGGGTTGTAG.....GCATCTCAC 15
::  |||||||  |||||||:  |||:  |||
83 al...ProLeucIyLeuGInGlyCysGluLeuProCysProAlaGluHis 98
14 ATT 12
::
99 Leu 99
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA12001.DAT:AAE03308

seq_documentation_block:

ID AAE03308 standard; Protein; 143 AA.

AC AAE03308;

DT 10-AUG-2001 (first entry)

DE Human gene 17 encoded secreted protein HTOI228, SEQ ID NO:82.

XX Human: secreted protein; proliferative disorder; cancer; tumour;
KW foetal abnormality; developmental abnormality; haematopoietic disorder;
KW immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
KW inflammation; allergy; neurological disorder; Alzheimer's disease;
KW Parkinson's disease; cognitive disorder; schizophrenia; asthma;
KW skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
KW cardiovascular disorder; angiodysplasia; kidney disorder;
KW gastrointestinal disorder; pregnancy-related disorder;
KW endocrine disorder; infection; wound healing; vulnerability;
KW cell culture; chemotaxis; food additive; gene therapy;
KW binding partner identification.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..31

FT Protein 32..132

XX /label= Human_mature_secretcd_protein

XX WC200134800-A1.

XX 17-MAY-2001.

XX 08-NOV-2000; 2000MO-US30674.

XX 12-NOV-1999; 99US-0164750.

XX 30-JUN-2000; 2000US-0215128.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Ruben SM, Komatsoulis GA, Ebner R, Fiscella M, Wei P;
XX
DR WPI: 2001-329085/34.
DR N-PSDB; AAD07721.

XX New nucleic acid molecules encoding human secreted proteins, used in
PT preventing, treating or ameliorating a disorder, e.g. Alzheimer's and
PT Parkinson's diseases and cancers -
XX
PS Claim 11; Page 467-468; 530pp; English.

XX AAD07705-AAD07759 represent cDNAs corresponding to 19 human secreted
CC protein genes, and AAE03292-AAE03346 represent the proteins they encode.
CC AAE03347-AAE03375 represent human secreted protein fragments or variants.
CC The genes and their secreted proteins are useful for preventing,
CC treating or ameliorating medical conditions, e.g., by protein or gene
CC therapy. Pathological conditions can be diagnosed by determining the
CC amount of the new protein in a sample or by determining the presence of
CC mutations in the new genes. Specific uses are described for each of the
CC 19 genes, based on the tissues in which they are most highly expressed,
CC and include developing products for the diagnosis or treatment of
CC proliferative disorders, cancer, tumours, foetal and developmental
CC abnormalities, haematopoietic disorders, diseases of the immune system,
CC AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,
CC allergies, neurological disorders (e.g., Alzheimer's disease,
CC Parkinson's disease), cognitive disorders, schizophrenia, asthma,
CC skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,
CC cardiovascular disorders, angiodysplasia, kidney disorders,
CC gastrointestinal disorders, pregnancy-related disorders, endocrine
CC disorders, and infections. The proteins can also be used to aid wound
CC healing and epithelial cell proliferation, to prevent skin aging due to
CC sunburn, to maintain organs before transplantation, for supporting cell
CC culture of primary tissues, to regenerate tissues, to identify their
CC cognate ligands or binding partners, and in chemotaxis, and can be used
CC as a food additive or preservative to modify storage properties.
CC Antibodies specific for a protein of the invention can be used in
CC alleviating symptoms associated with the disorders mentioned above, and
CC in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked
CC immunosorbent assay (ELISA). The present sequence represents a human
CC secreted protein of the invention.

SQ Sequence 143 AA;

alignment_scores:
Quality: 53.50 Length: 51
Ratio: 2.058 Gaps: 4
Percent Similarity: 50.980 Percent Identity: 35.294

alignment_block:

US-09-684-016-48411/rev x AAE03308 ..

Align seg 1/1 to: AAE03308 from: 1 to: 143

```
116 TGGATTAATTCAGATAGTATCACCCTTACATGTGAAA...CTTACCAG 70
|||||  |||  |||:  |||:  |||  |||
50 TTPASNIleuLeuTyrSerValGlySerIysValSerGlyLeuCysSe 66
69 TCAGTCAGCAGC.....CCTCGCA 53
||:|||||
66 TAsncysSerLeuValProGlyValAlaIAlaHisThrCysAsnProLysV 83
52 TTGGCCCATTAAGTCATCAGGGTTGTAG.....GCATCTCAC 15
::  |||||||  |||||||:  |||:  |||
83 al...ProLeucIyLeuGInGlyCysGluLeuProCysProAlaGluHis 98
14 ATT 12
::
99 Leu 99
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqp/AA1999.DAT:AAV6128

seq_documentation_block:

```

ID      AAY36128 standard; Protein; 180 AA.
AC      AAY36128;
DT      13-SEP-1999 (first entry)
DE      Extended human secreted protein sequence, SEQ ID NO. 513.
KW      Secreted protein; human; cytokine; cellular proliferation; cell movement;
KW      cellular differentiation; immune system regulator; anti-inflammatory;
KW      haematopoiesis regulator; tissue growth regulator; tumour inhibitor;
KW      reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;
KW      genetic disease.
OS      Homo sapiens.
PN      WO9331236-A2.
PD      24-JUN-1999.
PF      17-DEC-1998; 98W0-IB02122.
PI      10-AUG-1998; 98US-0096116.
PR      17-DEC-1997; 97US-0069957.
PR      09-FEB-1998; 98US-0074121.
PR      13-APR-1998; 98US-0081563.
PT      (GEST ) GENSET.
XX      Bougueleret L, Duclert A, Dumas Milne Edwards J;
XX      WPI: 1999-385906/32.
XX      N-PSDB; AAX97812.
XX      New isolated human secreted proteins
XX      Claim 9; Page 447; 516pp; English.
XX      This sequence is encoded by an extended human secreted protein coding
XX      sequence of the invention. The secreted proteins can be used in treating
XX      or controlling a variety of human conditions. The secreted proteins may
XX      act as cytokines or may affect cellular proliferation or differentiation
XX      or may act as immune system regulators, haematopoiesis regulators, tissue
XX      growth regulators, regulators of reproductive hormones or cell movement
XX      or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or
XX      tumour inhibition activity. The DNAs can be used in forensic procedures
XX      to identify individuals or in diagnostic procedures to identify
XX      individuals having genetic diseases resulting from abnormal expression of
XX      the genes corresponding to the extended cDNAs. They are also useful for
XX      constructing a high resolution map of the human chromosomes. They can
XX      also be used for gene therapy to control or treat genetic diseases.
XX      Sequence      180 AA;
XX
XX      alignment_scores:
XX          Quality:      53.00      Length:      28
XX          Ratio:        2.650      Gaps:       0
XX      Percent Similarity: 71.429      Percent Identity: 42.857
XX
XX      alignment_block:
XX      US-09-684-016-48411 x AAY36128 ..
XX
XX      Align seg 1/1 to: AAY36128 from: 1 to: 180
XX
XX      38 GACCTAATGGCGCAATGCCAGGCGTGCACATGACTGGTAAGTTTCACATGT 87
XX      |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
XX      69 AspiIeThrAlaAspLeuGluAsnIlePheAspIrpAsnValIysGlnLe 85
XX
XX      88 AAGGTGATACTATTCTTCATTTATTCACACACAA 121
XX      : |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
XX      85 upheutryrLeuSerAlaGluIuTyrrSerThrLys 96

```

```

seq_name: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT: AAB24029
seq_documentation_block:
ID      AAB24029 standard; Protein: 180 AA.
XX
XX
AC      AAB24029;
XX
XX
DT      25-JAN-2001 (first entry)
XX
XX      Human PRO3567 protein sequence SEQ ID NO:27.
DE
XX      Human; tumour; diagnosis; neoplastic disease; proliferation; cancer;
KW      identification; tumorigenesis; anticancer; detection.
XX
XX      Homo sapiens.
OS
PN      WO200053750-A1.
XX
XX      14-SEP-2000.
PD
XX      02-DEC-1999; 99WO-US28551.
XX
XX      08-MAR-1999; 99WO-US05078.
PR      01-SEP-1999; 99WO-US20111.
PR      29-OCT-1999; 99US-0162506.
PR      30-NOV-1999; 99WO-US28313.
PR      01-DEC-1999; 99WO-US28634.
XX
XX      (GETH ) GENENTECH INC.
PA
XX
XX      Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WI;
PI      WPI: 2000-594320/56.
XX
XX      N-PSDB; AAC58111.
DR
XX
XX      Antibodies specific for PRO polypeptides, used to diagnose and inhibit
PT      the growth of tumors in mammals, and to identify inhibitors of PRO
PR      polypeptide activity or expression -
XX
XX
XX      Claim 61; Fig 18; 226pp; English.
PS
XX
XX      The present invention describes an antibody that binds to a human
CC      protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;
CC      PRO3434; PRO1927; PRO3567; PRO1295; PRO1293; PRO1303; PRO4344; PRO4354;
CC      PRO4397; PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has
CC      anticancer activity and can be used to diagnose tumours in mammals, by
CC      detecting complex formation when the antibody is contacted with test
CC      cells. Increased expression of genes encoding (I) can also be detected
CC      to diagnose tumours. Agents which inhibit the activity of (I),
CC      especially the antibodies, or an antisense oligonucleotide which
CC      hybridises to genes encoding (I), can be used to inhibit tumour growth,
CC      preferably by inducing cell death. Methods from the present invention
CC      can be used to identify compounds which inhibit the biological activity
CC      of (I). AAC58019 to AAC58102 represent PCR primers and hybridisation
CC      probes used in examples from the present invention for human PRO
CC      sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human
CC      PRO polynucleotide and protein sequences given in the exemplification of
CC      the present invention.
XX
XX
XX      Sequence 180 AA;
SO

```


CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)
 CC are used (i) to identify agents suitable for treatment of uterine or
 CC endometrial cancer; (ii) directly for treating these forms of cancer
 CC (including expression from gene therapy vectors) and (iii) for
 CC generation of specific antibodies. (A) are identified by assembling ESTs
 CC (expressed sequence tags) from a particular tissue type before comparison
 CC of expression patterns. This allows a significantly longer fragment of
 CC the gene to be revealed, so should reduce the number of failures
 CC associated with the fact that ESTs from different libraries may represent
 CC different parts of the same unknown gene, distorting the estimated
 CC frequency of occurrence in a particular tissue. AAY5941-Y60328 represent
 CC protein fragments encoded by the human endometrium tumour cDNA library
 CC derived EST fragments represented in AAZ41981-742121.
 XX
 SQ Sequence 66 AA;

alignment_scores:
 Quality: 51.00 Length: 30
 Ratio: 2.550 Gaps: 1
 Percent Similarity: 66.667 Percent Identity: 33.333

alignment_block:
 US-09-684-016-48411/rev x AAY60285 ..

Align seg 1/1 to: AAY60285 from: 1 to: 66

```

110 AAATTGAGATAGTATCATCTTACATGTC.....AACTTACCACTCA 67
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
32 LysIleSerLeuSerPheSerIleuHisIleMetPheIysPheGlnAArgH1 48
66 GTGCAGCCCTCCGATTCGCCCATTTAGCTCATCAGGCTTGT 27
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:
48 scysArgGluArgValArgProCysGlyGluLeuMetCys 61

```

seq_name: /SID52/gcdata/geneseq/geneseq/AA1997.DAT:AAW20131

seq_documentation_block:
 ID AAW20131 standard; Protein: 288 AA.
 AC AAW20131;
 XX
 DT 08-JUL-1997 (first entry)
 XX
 DE H. pylori cell envelope transporter protein 13726562.aa.
 XX
 KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
 KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
 KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope; ds.
 XX
 OS Helicobacter pylori.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 8 /label= unknown
 FT /note= "encoded by YAA"
 FT
 XX W09640893-A1.
 XX
 PD 19-DEC-1996.
 XX
 PF 06-JUN-1996; 96WO-US09122.
 XX
 PR 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Berglindh OT, Smith D, Mellgaerd BL;
 XX
 DR WPI: 1997-052306/05.
 DR N-PSDB: AAT67376.
 XX

PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX
 PS Claim 56: Page 356; 1481pp; English.
 XX
 CC The present sequence shows a Helicobacter pylori cell envelope
 CC transporter protein.
 CC This protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 XX
 SQ Sequence 288 AA;

alignment_scores:
 Quality: 51.00 Length: 31
 Ratio: 2.684 Gaps: 0
 Percent Similarity: 61.290 Percent Identity: 38.710

alignment_block:
 US-09-684-016-48411/rev x AAW20131 ..

Align seg 1/1 to: AAW20131 from: 1 to: 288

```

110 AAATTGAGATAGTATCATCTTACATGTCGAACTTACCACTGACGAGCGAC 61
   ||| |||:|||||:|||||:|||||:|||||:|||||:|||||:
156 LysSerIysHisProValLeuLysAsnPheAsnLeuThrIleGlnIysci 172
60 CCCTGCATTCGCCCATTTAGCTCATCAGGCTTGTAGGCATCT 18
   : |||:|||||:|||||:|||||:|||||:|||||:|||||:
172 yGlnIysIleAlaLeuIleGlyHisSerGlyCysGlyLysSer 186

```

seq_name: /SID52/gcdata/geneseq/geneseq/AA1997.DAT:AAW20619

seq_documentation_block:
 ID AAW20619 standard; Protein: 377 AA.
 AC AAW20619;
 XX
 DT 14-JUL-1997 (first entry)
 XX
 DE H. pylori transporter protein, 02ae1161lori1.
 XX
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacterium; life cycle; activator;
 KW bacteria; inhibitor; duodenal ulcer disease; chronic gastritis;
 KW diagnosis.
 XX
 OS Helicobacter pylori.
 XX
 PD W09640893-A1.
 XX
 PF 19-DEC-1996.
 XX
 PR 06-JUN-1996; 96WO-US09122.
 XX
 PR 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Berglindh OT, Smith D, Mellgaerd BL;
 XX

XX WPI: 1997-052306/05.
 DR N-PSDB: AAT67872.

XX Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter

XX PS Claim 56: Page 1045-46; 1481pp; English.

XX This sequence represents a H. pylori transporter protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds,
 CC useful as potential H. pylori life cycle activators or inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the bacterial
 CC DNA. The sequences were analysed for ORF of at least 180 nucleotides,
 CC and the predicted coding regions defined by computer evaluation. To
 CC identify likely H. pylori antigens for vaccine development, the amino
 CC acid sequences predicted from various ORF were analysed for significant
 CC homology to other known or exported membrane proteins. Having identified
 CC and determined the sequences of interest, particular regions can be
 CC isolated from H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.

XX Sequence 377 AA;

alignment_scores:
 Quality: 51.00 Length: 31
 Ratio: 2.684 Gaps: 0
 Percent Similarity: 61.290 Percent Identity: 38.710

alignment_block:
 US-09-684-016-48411/rev x AAW20619 ..

Align seg 1/1 to: AAW20619 from: 1 to: 377

```

110 AATGTGAGATGATGACCTTACATGCACTTACAGTACAGTGCAG 61
    ||| |||   ::   ::   ::||| |||   ::
174 LysSerLysHisProValLeuLysAsnPheAsnLeuThrIleGlnLysG1 190
    :  ::||| |||   ::||| |||   ::||| |||
60 CCCTGCGATTGCGCATTAGTGCATCAGGGTGTGAAGGCATCT 18
    :  ::||| |||   ::||| |||   ::||| |||
190 yGlnLysIleAlaLeuIleGlyHisSerGlyCysGlyLysSer 204

```

THIS PAGE BLANK (USPTO)

OM of: US-09-684-016-48411 to: Issued_Patents_AA: * out_format : pfs
 Date: Mar 19, 2002 9:31 AM
 About: Results were produced by the Gencore software, version 4.5,
 Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=framer+np.model -DEV=xlh
 -O=/cgn2_1/USPTO.spool/US09684016/runat.19032002.084440.20900/esp_query.fasta_1.182
 -De=Issued_Patents_AA -QFMT=fasta -SUFFIX=std.ra1
 -GAPOP=12.000 -GAPOP=4.500 -MINMATCH=0.100 -LOOPEL=0.000
 -LOOPEXT=0.000 -GAPOP=4.500 -GAPOP=10.000
 -XGAPOP=0.500 -XGAPOP=6.000 -XGAPOP=7.000 -XGAPOP=10.000
 -XGAPOP=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45 -DOCALLIGN=200
 -THR_SCORE=pcpt -THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL
 -OUTFMT=pfs -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
 -USER=US09684016.@CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPHY -WAIT -THREADS=1

Search information block:
 Query: US-09-684-016-48411
 Query length: 123
 Database: Issued_Patents_AA: *
 Database sequences: 212252
 Database length: 22503292
 Search time (sec): 25.790000

score list:

Seqname	Strd Orig	zScore	EScore	Len	Documentation	
/cgn2_6/p/ptodata/2/1aa/6A.COMB.pep:US-08-762-428A-6		56.00	118.65	2.89	1288	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-09-085-199B-4		54.00	116.51	5.35	914	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-09-085-199B-5		54.00	114.85	5.55	1090	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-549-515-9		51.00	122.45	11.03	207	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-866-545-4		50.00	129.21	12.63	76	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-332-087A-11		50.00	122.25	14.73	159	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-219-237B-6		50.00	122.25	14.73	159	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-477-347-15		50.00	122.25	14.73	159	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-476-862-6		50.00	122.25	14.73	159	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-468-560C-6		50.00	122.25	14.73	159	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-974-022-50		50.00	119.02	15.82	224	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-795-447A-50		50.00	119.02	15.82	224	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-974-186-50		50.00	119.02	15.82	224	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-795-446B-50		50.00	119.02	15.82	224	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-864-804-1		50.00	113.10	18.04	420	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-09-041-886-2		50.00	112.94	18.10	427	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-09-006-353A-5		50.00	112.94	18.10	427	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-553-436A-8		50.00	107.43	20.45	766	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-453-472-6		48.50	104.08	33.77	713	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-038-948-7		48.50	104.08	33.77	713	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-038-948-8		48.50	104.08	33.77	713	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-038-948-10		48.50	104.08	33.77	713	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-453-952-6		48.50	104.08	33.77	713	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-862-903-6		48.50	104.08	33.77	713	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-549-515-6		48.00	114.49	30.93	205	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-549-515-12		48.00	114.49	30.93	205	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-249-112-4		48.00	106.33	37.05	487	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-249-112-4		48.00	106.33	37.05	487	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-09-211-704A-7		48.00	105.73	37.35	519	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-09-115-954-4		48.00	102.73	40.13	714	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-357-598-12		48.00	101.13	41.57	846	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-09-003-289-12		48.00	101.13	41.57	846	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-09-003-289-12		48.00	101.13	41.57	846	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-09-115-954-8		48.00	99.25	43.34	1032	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-09-115-954-8		48.00	99.25	43.34	1032	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-09-352-990-18		47.00	105.50	50.18	400	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-445-586-9		47.00	101.95	54.28	583	1
/cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-08-484-493-14		47.00	101.95	54.28	583	1

```

seq_name: /cgn2_6/p/ptodata/2/1aa/6A.COMB.pep:US-08-762-428A-6
seq_documentation_block:
: Sequence 6, Application US/08762428A
: Patent No. 6120993
: GENERAL INFORMATION:
: APPLICANT: Ye, Guo-Jie
: APPLICANT: Breslow, Esther M.
: APPLICANT: Meister, Alton
: TITLE OF INVENTION: 5-OXOPROLINASE
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: NIXON, HARGRAVE, DEVANS & DOYLE LLP
: STREET: Clinton Square, P.O. Box 1051
: CITY: Rochester
: STATE: New York
: COUNTRY: USA
: ZIP: 14603
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/762,428A
: FILING DATE:
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: TIMIAN, SUSAN J.
: REGISTRATION NUMBER: 34,103
: REFERENCE/DOCKET NUMBER: 19603/1330
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 716-263-1636
: TELEFAX: 716-263-1600
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1288 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-762-428A-6

alignment_scores:
Quality: 56.00 Length: 30
Ratio: 2.800 Gaps: 1
Percent Similarity: 66.667 Percent Identity: 40.000

alignment_block:
US-09-684-016-48411/rev x US-08-762-428A-6 ..
Align seg 1/1 to: US-08-762-428A-6 from: 1 to: 1288
86 CATGGAACATACAGTCAGTCAGCCCT.....CGCATGCCCAT 43
|||||.....|||||
231 gGlyHisThrAlaCysAlaAspAlaTyrLeuThrProThr 244
|||||.....|||||
215 HisValSerLeuSerGluValMetProMetValArgIleValProAr 231
|||||.....|||||
42 AGTCATCAGGGTGTGTAAGCATTCATCATTTACATAC 3
|||||.....|||||
seq_name: /cgn2_6/p/ptodata/2/1aa/6B.COMB.pep:US-09-085-199B-4
seq_documentation_block:
: Sequence 4, Application US/09085199B
: Patent No. 6235879

```

```
seq_documentation_block:
; Sequence 9, Application US/08549515
; Patent No. 6054123
; GENERAL INFORMATION:
```



```

; APPLICANT: Loosmore, Sheena M
; APPLICANT: Klein, Michael H
; TITLE OF INVENTION: Haemophilus Influenzae
; TITLE OF INVENTION: Dimethylsulphoxide Reductase Enzyme
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/549,515
; FILING DATE: 27-OCT-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-522
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 207 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-549-515-9

alignment_scores:
      Quality: 51.00      Length: 29
      Ratio: 3.400      Gaps: 1
      Percent Similarity: 51.724      Percent Identity: 34.483

alignment_block:
US-09-684-016-48411 x US-08-549-515-9  ..
Align seg 1/1 to: US-08-549-515-9 from: 1 to: 207

      2 AGGATATGTAAATGTGAGATGCTTACACCGCTGATGACCTA..... 43
      |||||||      |||||||      ::
      103 ArgTyrCysHsMetAlaCysProTyrGlyAlaProGlnTyrAsnGluTh 119
      44 .....ATGGCGCAATGCCAGGCGCTGACTGAC 70
      |||:|||||:|||||:|||||:||||| 111
      119 rLysGlnHsMetHrLysCysAspGlyCysTyrAsp 131

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-866-545-4

seq_documentation_block:
; Sequence 4, Application US/08866545
; Patent No. 6265535
; GENERAL INFORMATION:
; APPLICANT: Greene, Mark I.
; APPLICANT: Murali, Ramachandran
; APPLICANT: Takasaki, Wataru
; TITLE OF INVENTION: PEPTIDES AND PEPTIDE
; TITLE OF INVENTION: ANALOGUES DESIGNED FROM BINDING SITES OF TUMOR
; TITLE OF INVENTION: NECROSIS FACTOR RECEPTOR SUPERFAMILY AND THEIR
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York

```

```

; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/866,545
; FILING DATE: 30-MAY-1997
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Cortuzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 009113-0004-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNTE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 76 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6265535e
;
US-08-866-545-4

alignment_scores:
      Quality: 50.00      Length: 19
      Ratio: 4.167      Gaps: 0
      Percent Similarity: 63.158      Percent Identity: 36.842

alignment_block:
US-09-684-016-48411 x US-08-866-545-4  ..
Align seg 1/1 to: US-08-866-545-4 from: 1 to: 76

      8 TGTAAATGTGAGATGCTTACACCGCTGATGACCTAATGGCGCAATGCGA 57
      |||:|||||      |||:|||||      :|||:|||||
      44 CysArgCysAlaTyrGlyTyrTyrGlnAspGluThrGlyArgCysG1 60
      58 GGGCTGC 64
      ||:|||||
      60 uAlaCys 62

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-232-087A-11

seq_documentation_block:
; Sequence 11, Application US/08232087A
; Patent No. 5866372
; GENERAL INFORMATION:
; APPLICANT: Stein, Harald
; APPLICANT: Dikop, Horst
; APPLICANT: Latza, Ute
; TITLE OF INVENTION: Lymphoid CD30-Antigen
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: 8110 Gatehouse Road, Suite 500 East
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22042
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```

```

: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/232,087A
: FILING DATE: 08-SEP-1994
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Murphy Jr., Gerald M.
: REGISTRATION NUMBER: 28,977
: REFERENCE/DOCKET NUMBER: 736-103P
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703) 205-8000
: TELEFAX: (703) 205-8050
: TELEX: 248345
: INFORMATION FOR SEQ ID NO: 11:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 159 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: FRAGMENT TYPE: Internal
: FEATURE:
: NAME/KEY: Protein
: LOCATION: 1..159
: OTHER INFORMATION: /note="NGFR, see Fig. 5"
:
: US-08-232-087A-11

alignment_scores:
      Quality: 50.00      Length: 20
      Ratio: 3.846      Gaps: 0
      Percent Similarity: 65.000      Percent Identity: 35.000

alignment_block:
US-09-684-016-48411 x US-08-232-087A-11 ..

Align seg 1/1 to: US-08-232-087A-11 from: 1 to: 159

      8 TGTAAATGTGAGATGCTTACACCTGATGACCTAATGCGCAATGCGA 57
      |||::||| ||| |||::|||
      77 CysArgCysAlaTyrGlyTyrTyrGlnAspGluThrThcGlyArgCysG1 93
      58 GGGCTGCGACT 67
      ||::|||::
      93 uAlaCysSer 96

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-219-237B-6

seq_documentation_block:
: Sequence 6, Application US/08219237B
: Patent No. 5874546
: GENERAL INFORMATION:
: APPLICANT: NAGATA, Shigekazu
: APPLICANT: ITOH, Naoto
: APPLICANT: YONEHARA, Shin
: TITLE OF INVENTION: DNA Coding for Human Cell Surface Antigen
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: James W. Hellwege
: STREET: P.O. Box 2266 Eads Station
: CITY: Arlington
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/219,237B
: FILING DATE: 28-MAR-1994

```

```

: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/872,129
: FILING DATE: 22-APR-1992
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: James W. Hellwege
: REGISTRATION NUMBER: 28,808
: REFERENCE/DOCKET NUMBER: 516762
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 159 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
:
: US-08-219-237B-6

alignment_scores:
      Quality: 50.00      Length: 19
      Ratio: 4.167      Gaps: 0
      Percent Similarity: 63.158      Percent Identity: 36.842

alignment_block:
US-09-684-016-48411 x US-08-219-237B-6 ..

Align seg 1/1 to: US-08-219-237B-6 from: 1 to: 159

      8 TGTAAATGTGAGATGCTTACACCTGATGACCTAATGCGCAATGCGA 57
      |||::||| ||| |||::|||
      77 CysArgCysAlaTyrGlyTyrTyrGlnAspGluThrThcGlyArgCysG1 93
      58 GGGCTGCG 64
      ||::|||
      93 uAlaCys 95

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-477-347-15

seq_documentation_block:
: Sequence 15, Application US/08477347
: Patent No. 6232446
: GENERAL INFORMATION:
: APPLICANT: WALLACH, David
: APPLICANT: BIGDA, Jacek
: APPLICANT: BELETSKY, Igor
: APPLICANT: METT, Igor
: TITLE OF INVENTION: TNF LIGANDS
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BROWDY AND NEIMARK
: STREET: 419 Seventh Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20004
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/477,347
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/115,685
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: IL 106271
: FILING DATE: 08-JUL-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Townsend, G. Kevin
: REGISTRATION NUMBER: 34,033

```

```
;
; REFERENCE/DOCKET NUMBER: WALLACH-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 159 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-477-347-15

alignment_scores:
      Quality: 50.00      Length: 19
      Ratio: 4.167      Gaps: 0
Percent Similarity: 63.158      Percent Identity: 36.842

alignment_block:
US-09-684-016-48411 x US-08-477-347-15      ..

Align seg 1/1 to: US-08-477-347-15 from: 1 to: 159

      8 TGTAAATGTGAGATGCTTACACCCGTGATGACCTAATGGCGCAATGCGA 57
      |||::||| ||| |||::|
      77 CysatgCysalaTyrglyTyrtYrglnspsluThrThrglyArgCysgl 93
      58 GGGCTGC 64
      |::|||
      93 uAlacys 95

seq_name: /cgn2_6/ptodata/2/1aa/6b_COMB.pep:US-08-476-862-6

seq_documentation_block:
; Sequence 6, Application US/08476862
; Patent No. 6262239
; GENERAL INFORMATION:
; APPLICANT: WALLACH, David
; APPLICANT: BIGDA, Jacek
; APPLICANT: BELETSKY, Igor
; APPLICANT: METT, Igor
; APPLICANT: ENGELMANN, Hartmut
; TITLE OF INVENTION: TNF INHIBITORS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,862
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 107267
; FILING DATE: 12-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 94039
; FILING DATE: 06-APR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 91229
; FILING DATE: 06-AUG-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: IL 90339
```

```
;
; FILING DATE: 18-MAY-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWDY, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: WALLACH-12A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 159 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-476-862-6

alignment_scores:
      Quality: 50.00      Length: 19
      Ratio: 4.167      Gaps: 0
Percent Similarity: 63.158      Percent Identity: 36.842

alignment_block:
US-09-684-016-48411 x US-08-476-862-6      ..

Align seg 1/1 to: US-08-476-862-6 from: 1 to: 159

      8 TGTAAATGTGAGATGCTTACACCCGTGATGACCTAATGGCGCAATGCGA 57
      |||::||| ||| |||::|
      77 CysatgCysalaTyrglyTyrtYrglnspsluThrThrglyArgCysgl 93
      58 GGGCTGC 64
      |::|||
      93 uAlacys 95

seq_name: /cgn2_6/ptodata/2/1aa/6b_COMB.pep:US-08-468-560C-6

seq_documentation_block:
; Sequence 6, Application US/08468560C
; Patent No. 6270998
; GENERAL INFORMATION:
; APPLICANT: ITOH, Naoto
; APPLICANT: NAGATA, Shigekazu
; APPLICANT: YONEMURA, Shin
; TITLE OF INVENTION: DNA CODING FOR HUMAN CELL SURFACE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH, LLP.
; STREET: P.O. BOX 747
; CITY: FALLS CHURCH
; STATE: VA
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,560C
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MORPHY JR., GERLAD M.
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 20-4393P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-205-8000
; TELEFAX: 703-205-8050
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
```

LENGTH: 159 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-468-560C-6

alignment_scores:
Quality: 50.00 Length: 19
Ratio: 4.167 Gaps: 0
Percent Similarity: 63.158 Percent Identity: 36.842

alignment_block:
US-09-684-016-48411 x US-08-468-560C-6 ..

Align seg 1/1 to: US-08-468-560C-6 from: 1 to: 159

8 TGTAAATGTGAGATGCCTTACACCCCTGATGACCTAATGGCGCAATGCCA 57
||||:|||||
77 CysArgCysAlaTyrGlyTyrGlnAspGluThrThrglyArgCysG1 93
58 GGGCTGC 64
|:::|
93 uAlaCys 95

seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-974-022-50

seq_documentation_block:

Sequence 50, Application US/08974022
Patent No. 6015938
GENERAL INFORMATION:
APPLICANT: Boyle, William J.
APPLICANT: Lacey, David L.
APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: OSTEOPROTEGERIN
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974.022
FILING DATE: 12-DEC-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/577,788
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-378
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 224 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-974-022-50

alignment_scores:
Quality: 50.00 Length: 19
Ratio: 4.167 Gaps: 0

Percent Similarity: 63.158 Percent Identity: 36.842

alignment_block:

US-09-684-016-48411 x US-08-974-022-50 ..

Align seg 1/1 to: US-08-974-022-50 from: 1 to: 224

8 TGTAAATGTGAGATGCCTTACACCCCTGATGACCTAATGGCGCAATGCCA 57
||||:|||||
107 CysArgCysAlaTyrGlyTyrGlnAspGluThrThrglyArgCysG1 123
58 GGGCTGC 64
|:::|
123 uAlaCys 125

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-795-445A-50

seq_documentation_block:

Sequence 50, Application US/08795445A
Patent No. 6284485
GENERAL INFORMATION:
APPLICANT: Boyle, William J.
APPLICANT: Lacey, David L.
APPLICANT: Calzone, Frank J.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: OSTEOPROTEGERIN
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795.445A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/577,788
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-378
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 224 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-795-445A-50

alignment_scores:
Quality: 50.00 Length: 19
Ratio: 4.167 Gaps: 0
Percent Similarity: 63.158 Percent Identity: 36.842

alignment_block:

US-09-684-016-48411 x US-08-795-445A-50 ..

Align seg 1/1 to: US-08-795-445A-50 from: 1 to: 224

8 TGTAAATGTGAGATGCCTTACACCCCTGATGACCTAATGGCGCAATGCCA 57
||||:|||||
107 CysArgCysAlaTyrGlyTyrGlnAspGluThrThrglyArgCysG1 123

```

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-795-447A-50
seq_documentation_block:
  Sequence 50, Application US/08795447A
  Patent No. 6284728
  GENERAL INFORMATION:
    APPLICANT: Boyle, William J.
    APPLICANT: Lacey, David L.
    APPLICANT: Calzone, Frank J.
    TITLE OF INVENTION: Osteoprotegerin
    NUMBER OF SEQUENCES: 53
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Amgen Inc.
      STREET: One Amgen Center Drive
      CITY: Thousand Oaks
      STATE: California
      COUNTRY: USA
      ZIP: 91320-1789
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    OPERATING SYSTEM: IBM PC compatible
    SOFTWARE: Patentin Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/08/795,447A
      FILING DATE:
      CLASSIFICATION: 514
      ATTORNEY/AGENT INFORMATION:
        NAME: Winter, Robert B.
        REFERENCE/DOCKET NUMBER: A-378D2
    INFORMATION FOR SEQ ID NO: 50:
      SEQUENCE CHARACTERISTICS:
        LENGTH: 224 amino acids
        TYPE: amino acid
        STRANDEDNESS: single
        TOPOLOGY: linear
        MOLECULE TYPE: protein
      US-08-795-447A-50

  alignment_scores:
    Quality: 50.00      Length: 19
    Ratio: 4.167        Gaps: 0
    Percent Similarity: 63.158   Percent Identity: 36.842

  alignment_block:
    US-09-684-016-48411 x US-08-795-447A-50
  Align seg 1/1 to: US-08-795-447A-50 from: 1 to: 224

  seq_documentation_block:
    Sequence 50, Application US/08974186
    Patent No. 6284740
    GENERAL INFORMATION:
      APPLICANT: Boyle, William J.
      APPLICANT: Lacey, David L.
      APPLICANT: Calzone, Frank J.
      APPLICANT: Chang, Ming-Shi
      TITLE OF INVENTION: Osteoprotegerin
      NUMBER OF SEQUENCES: 53
      CORRESPONDENCE ADDRESS:
        ADDRESSEE: Amgen Inc.
        STREET: 1840 Dehavilland Drive
        CITY: Thousand Oaks
        STATE: California
        COUNTRY: USA
        ZIP: 91320-1789
      COMPUTER READABLE FORM:
        MEDIUM TYPE: Floppy disk

  seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-974-186-50
seq_documentation_block:
  Sequence 50, Application US/08795446B
  Patent No. 6288032
  GENERAL INFORMATION:
    APPLICANT: Boyle, William J.
    APPLICANT: Lacey, David L.
    APPLICANT: Calzone, Frank J.
    APPLICANT: Chang, Ming-Shi
    TITLE OF INVENTION: OSTEOPROTEGERIN
    NUMBER OF SEQUENCES: 53
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Amgen Inc.
      STREET: 1840 Dehavilland Drive
      CITY: Thousand Oaks
      STATE: California
      COUNTRY: USA
      ZIP: 91320-1789
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk

  seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-795-446B-50
seq_documentation_block:
  Sequence 50, Application US/08795446B
  Patent No. 6288032
  GENERAL INFORMATION:
    APPLICANT: Boyle, William J.
    APPLICANT: Lacey, David L.
    APPLICANT: Calzone, Frank J.
    APPLICANT: Chang, Ming-Shi
    TITLE OF INVENTION: OSTEOPROTEGERIN
    NUMBER OF SEQUENCES: 53
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Amgen Inc.
      STREET: 1840 Dehavilland Drive
      CITY: Thousand Oaks
      STATE: California
      COUNTRY: USA
      ZIP: 91320-1789
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk

  seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-974-186-50
seq_documentation_block:
  Sequence 50, Application US/08795446B
  Patent No. 6288032
  GENERAL INFORMATION:
    APPLICANT: Boyle, William J.
    APPLICANT: Lacey, David L.
    APPLICANT: Calzone, Frank J.
    APPLICANT: Chang, Ming-Shi
    TITLE OF INVENTION: OSTEOPROTEGERIN
    NUMBER OF SEQUENCES: 53
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Amgen Inc.
      STREET: 1840 Dehavilland Drive
      CITY: Thousand Oaks
      STATE: California
      COUNTRY: USA
      ZIP: 91320-1789
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk

  seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-974-186-50
seq_documentation_block:
  Sequence 50, Application US/08795446B
  Patent No. 6288032
  GENERAL INFORMATION:
    APPLICANT: Boyle, William J.
    APPLICANT: Lacey, David L.
    APPLICANT: Calzone, Frank J.
    APPLICANT: Chang, Ming-Shi
    TITLE OF INVENTION: OSTEOPROTEGERIN
    NUMBER OF SEQUENCES: 53
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Amgen Inc.
      STREET: 1840 Dehavilland Drive
      CITY: Thousand Oaks
      STATE: California
      COUNTRY: USA
      ZIP: 91320-1789
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk

  seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-974-186-50
seq_documentation_block:
  Sequence 50, Application US/08795446B
  Patent No. 6288032
  GENERAL INFORMATION:
    APPLICANT: Boyle, William J.
    APPLICANT: Lacey, David L.
    APPLICANT: Calzone, Frank J.
    APPLICANT: Chang, Ming-Shi
    TITLE OF INVENTION: OSTEOPROTEGERIN
    NUMBER OF SEQUENCES: 53
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Amgen Inc.
      STREET: 1840 Dehavilland Drive
      CITY: Thousand Oaks
      STATE: California
      COUNTRY: USA
      ZIP: 91320-1789
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk

  seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-974-186-50
seq_documentation_block:
  Sequence 50, Application US/08795446B
  Patent No. 6288032
  GENERAL INFORMATION:
    APPLICANT: Boyle, William J.
    APPLICANT: Lacey, David L.
    APPLICANT: Calzone, Frank J.
    APPLICANT: Chang, Ming-Shi
    TITLE OF INVENTION: OSTEOPROTEGERIN
    NUMBER OF SEQUENCES: 53
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Amgen Inc.
      STREET: 1840 Dehavilland Drive
      CITY: Thousand Oaks
      STATE: California
      COUNTRY: USA
      ZIP: 91320-1789
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk

  seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-974-186-50
seq_documentation_block:
  Sequence 50, Application US/08795446B
  Patent No. 6288032
  GENERAL INFORMATION:
    APPLICANT: Boyle, William J.
    APPLICANT: Lacey, David L.
    APPLICANT: Calzone, Frank J.
    APPLICANT: Chang, Ming-Shi
    TITLE OF INVENTION: OSTEOPROTEGERIN
    NUMBER OF SEQUENCES: 53
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Amgen Inc.
      STREET: 1840 Dehavilland Drive
      CITY: Thousand Oaks
      STATE: California
      COUNTRY: USA
      ZIP: 91320-1789
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk

  seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-974-186-50
seq_documentation_block:
  Sequence 50, Application US/08795446B
  Patent No. 6288032
  GENERAL INFORMATION:
    APPLICANT: Boyle, William J.
    APPLICANT: Lacey, David L.
    APPLICANT: Calzone, Frank J.
    APPLICANT: Chang, Ming-Shi
    TITLE OF INVENTION: OSTEOPROTEGERIN
    NUMBER OF SEQUENCES: 53
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Amgen Inc.
      STREET: 1840 Dehavilland Drive
      CITY: Thousand Oaks
      STATE: California
      COUNTRY: USA
      ZIP: 91320-1789
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk

  seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-974-186-50
seq_documentation_block:
  Sequence 50, Application US/08795446B
  Patent No. 6288032
  GENERAL INFORMATION:
    APPLICANT: Boyle, William J.
    APPLICANT: Lacey, David L.
    APPLICANT: Calzone, Frank J.
    APPLICANT: Chang, Ming-Shi
    TITLE OF INVENTION: OSTEOPROTEGERIN
    NUMBER OF SEQUENCES: 53
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Amgen Inc.
      STREET: 1840 Dehavilland Drive
      CITY: Thousand Oaks
      STATE: California
      COUNTRY: USA
      ZIP: 91320-1789
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk

  seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-974-186-50
seq_documentation_block:
  Sequence 50, Application US/08795446B
  Patent No. 6288032
  GENERAL INFORMATION:
    APPLICANT: Boyle, William J.
    APPLICANT: Lacey, David L.
    APPLICANT: Calzone, Frank J.
    APPLICANT: Chang, Ming-Shi
    TITLE OF INVENTION: OSTEOPROTEGERIN
    NUMBER OF SEQUENCES: 53
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Amgen Inc.
      STREET: 1840 Dehavilland Drive
      CITY: Thousand Oaks
      STATE: California
      COUNTRY: USA
      ZIP: 91320-1789
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk

  seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-974-186-50
seq_documentation_block:
  Sequence 50, Application US/08795446B
  Patent No. 6288032
  GENERAL INFORMATION:
    APPLICANT: Boyle, William J.
    APPLICANT: Lacey, David L.
    APPLICANT: Calzone, Frank J.
    APPLICANT: Chang, Ming-Shi
    TITLE OF INVENTION: OSTEOPROTEGERIN
    NUMBER OF SEQUENCES: 53
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Amgen Inc.
      STREET: 1840 Dehavilland Drive
      CITY: Thousand Oaks
      STATE: California
      COUNTRY: USA
      ZIP: 91320-1789
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk

  seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-974-186-50
seq_documentation_block:
  Sequence 50, Application US/08795446B
  Patent No. 6288032
  GENERAL INFORMATION:
    APPLICANT: Boyle, William J.
    APPLICANT: Lacey, David L.
    APPLICANT: Calzone, Frank J.
    APPLICANT: Chang, Ming-Shi
    TITLE OF INVENTION: OSTEOPROTEGERIN
    NUMBER OF SEQUENCES: 53
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Amgen Inc.
      STREET: 1840 Dehavilland Drive
      CITY: Thousand Oaks
      STATE: California
      COUNTRY: USA
      ZIP: 91320-1789
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk

  seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-974-186-50
seq_documentation_block:
  Sequence 50, Application US/08795446B
  Patent No. 6288032
  GENERAL INFORMATION:
    APPLICANT: Boyle, William J.
    APPLICANT: Lacey, David L.
    APPLICANT: Calzone, Frank J.
    APPLICANT: Chang, Ming-Shi
    TITLE OF INVENTION: OSTEOPROTEGERIN
    NUMBER OF SEQUENCES: 53
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Amgen Inc.
      STREET: 1840 Dehavilland Drive
      CITY: Thousand Oaks
      STATE: California
      COUNTRY: USA
      ZIP: 91320-1789
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk

  seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-974-186-50
seq_documentation_block:
  Sequence 50, Application US/08795446B
  Patent No. 6288032
  GENERAL INFORMATION:
    APPLICANT: Boyle, William J.
    APPLICANT: Lacey, David L.
    APPLICANT: Calzone, Frank J.
    APPLICANT: Chang, Ming-Shi
    TITLE OF INVENTION: OSTEOPROTEGERIN
    NUMBER OF SEQUENCES: 53
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: Amgen Inc.
      STREET: 1840 Dehavilland Drive
      CITY: Thousand Oaks
      STATE: California
      COUNTRY: USA
      ZIP: 91320-1789
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk

  seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-08-974-186-50
seq_documentation_block:
  Sequence 50, Application US/0879
```

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,446B
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/577,788
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-378
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 224 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-795-446B-50

```

```

alignment_scores:
Quality: 50.00 Length: 19
Ratio: 4.167 Gaps: 0
Percent Similarity: 63.158 Percent Identity: 36.842

```

alignment_block:

```
US-09-684-016-48411 x US-08-795-446B-50 ..
```

```
Align seg 1/1 to: US-08-795-446B-50 from: 1 to: 224
```

```

8 TGTAAATGTGATGCTTACAAACCTGATGACCTATGCGCGCATGCGA 57
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
107 CysatGcysAlatYrGlyTyrGlnaspGluThrGlyArgcysgl 123
58 GGGCTGC 64
123 uAlaCys 125

```

```
seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-864-804-1
```

```

seq_documentation_block:
; Sequence 1, Application US/08864804
; Patent No. 5948641
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purni
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: METAL RESPONSE ELEMENT BINDING PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/864,804
; FILING DATE: Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.

```

```

REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0307 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 420 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LIVERP02
CLONE: 2048959
US-08-864-804-1

```

```

alignment_scores:
Quality: 50.00 Length: 23
Ratio: 3.571 Gaps: 0
Percent Similarity: 60.870 Percent Identity: 39.130

```

alignment_block:

```
US-09-684-016-48411 x US-08-864-804-1 ..
```

```
Align seg 1/1 to: US-08-864-804-1 from: 1 to: 420
```

```

5 TATTGTAATGTGATGCTTACAAACCTGATGACCTATGCGCGCATG 54
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
189 TyrcystyrCysglYgIyProglYgluTrpaspLeuLysMetLeuGlnCy 205
55 CGAGGGCTGCACCTGACTG 73
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
205 sarGserCysLeuGlnTrp 211

```

```
seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-086-483A-4
```

```

seq_documentation_block:
; Sequence 4, Application US/09086483A
; Patent No. 6214580
; GENERAL INFORMATION:
; APPLICANT: NI, et al.
; TITLE OF INVENTION: HUMAN TUMOR NECROSIS FACTOR RECEPTOR TR10
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE
; STATE: MD
; COUNTRY: US
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/086,483A
; FILING DATE: May-29-98
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/050,936
; FILING DATE: May-30-97
; CLASSIFICATION: 435
; APPLICATION NUMBER: 60/069,112
; FILING DATE: Dec-9-97
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BROOKES, ANDERS A.
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PF379

```


8 TGTAAATGAGATGCTTACCAACCTGATGACCTAATGGCGCATCGCA 57
||||:|||||
107 CysATGcGcYalATrGlyTYrTYrGlnAspLutHrThrGlyAbgCysG1 123
58 GGGCTGC 64
|:|||||
123 uAlAcys 125

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-553-436A-8

seq_documentation_block:

; Sequence 8, Application US/08553436A
; Patent No. 5866790
; GENERAL INFORMATION:
; APPLICANT: HESSE, Holger
; APPLICANT: MULLER-ROBER, Bernd
; TITLE OF INVENTION: DNA SEQUENCES AND PLASMIDS FOR THE
; TITLE OF INVENTION: PREPARATION OF SUGAR BEET WITH CHANGED SUCROSE
; TITLE OF INVENTION: CONCENTRATION
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/553.436A
; FILING DATE: 17-NOV-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/EP94/01671
; FILING DATE: 20-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 4317596.1
; FILING DATE: 24-MAY-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meilman, Edward
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/951-117
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 382-0700
; TELEFAX: (212) 382-0888
; TELEX: 236925
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 766 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-553-436A-8

alignment_scores:

Quality:	50.00	Length:	57
Ratio:	2.273	Gaps:	2
Percent Similarity:	38.596	Percent Identity:	26.316

alignment_block:

US-09-684-016-48411/rev x US-08-553-436A-8 ..

Align seg 1/1 to: US-08-553-436A-8 from: 1 to: 766

105 GAAGATAAGTATCACCCTTACATGTGAA..... 79
|||||
399 GluAspLysTYrHisPheSerCysGlnPheSerAlaAspLeuMetAlaLe 415

78ACTTACCACTCAGT..... 64
|||||
415 tAsnHisAlaAspPheIleIleHrSerThrTYrGlnGluIleAlaGlyT 432
63CAGCCCTGCGATTCGCCCATTAGTCATCA 34
432 hTrysAsnThrValGlyGlnTYrGlnSerHisLysAlaPheHrPhePro 448
33 GGGTTGTAGGACATTCACAT 13
|||||
449 GlyLeuTYrArgValValHis 455

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-453-472-6

seq_documentation_block:

; Sequence 6, Application US/08453472
; Patent No. 5626846
; GENERAL INFORMATION:
; APPLICANT: DEAN, JURRIEN
; TITLE OF INVENTION: CONTRACEPTIVE VACCINE
; TITLE OF INVENTION: BASED ON ALLOIMMUNIZATION WITH ZONA PELLUCIDA
; TITLE OF INVENTION: POLYPEPTIDES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453.472
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/038,948
; FILING DATE: 26-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,462
; FILING DATE: 20-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/364,379
; FILING DATE: 12-JUN-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AOTH
; REGISTRATION NUMBER: 36,434
; REFERENCE/DOCKET NUMBER: 2026-4032 US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 713
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: mouse
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL TYPE:
; CELL LINE:

ORGANELLE:
FEATURE:
NAME/KEY: ZP2
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: mouse ZP2 protein
US-08-453-472-6

alignment_scores:
Quality: 48.50 Length: 38
Ratio: 2.553 Gaps: 2
Percent Similarity: 50.000 Percent Identity: 34.211

alignment_block:
US-09-684-016-48411 x US-08-453-472-6 ..

Align seg 1/1 to: US-08-453-472-6 from: 1 to: 713

8 TGTAAATGTGAGATGCTTACAACTGATGACCTAATGGCGCAATGCGA 57
||| ||||| ||| ::| ||||| ||||| ::
353 CysHisScysGluSerProValSerIleAspGluLeucCysAlaGln...As 368
58 GGGCTGCACCTGACTGTGAAGTTTCACATGTAAAGTGATCTTACTTCA 107
||| ||||| ::|
368 pGlyPheMetAspPhe.....GluV 375
108 TTTATTCACACAA 121
::| ||||| |||
375 alTySerHisGln 379

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-038-948-7

seq_documentation_block:

Sequence 7, Application US/08038948
Patent No. 5641487
GENERAL INFORMATION:
APPLICANT: DEAN, JURRIEN
TITLE OF INVENTION: CONTRACEPTIVE VACCINE BASED ON
NUMBER OF SEQUENCES: 14
TITLE OF INVENTION: ALLOIMMUNIZATION WITH ZONA PELLUCIDA POLYPEPTIDES
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/038,948
FILING DATE: 26-MAR-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/930,462
FILING DATE: 20-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/364,379
FILING DATE: 12-JUN-1989
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, Watson T.
REGISTRATION NUMBER: 26,581
REFERENCE/DOCKET NUMBER: 99152/E-266-88/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:

LENGTH: 713 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-038-948-7

alignment_scores:
Quality: 48.50 Length: 38
Ratio: 2.553 Gaps: 2
Percent Similarity: 50.000 Percent Identity: 34.211

alignment_block:
US-09-684-016-48411 x US-08-038-948-7 ..

Align seg 1/1 to: US-08-038-948-7 from: 1 to: 713

8 TGTAAATGTGAGATGCTTACAACTGATGACCTAATGGCGCAATGCGA 57
||| ||||| ||| ::| ||||| ||||| ::
353 CysHisScysGluSerProValSerIleAspGluLeucCysAlaGln...As 368
58 GGGCTGCACCTGACTGTGAAGTTTCACATGTAAAGTGATCTTACTTCA 107
||| ||||| ::|
368 pGlyPheMetAspPhe.....GluV 375
108 TTTATTCACACAA 121
::| ||||| |||
375 alTySerHisGln 379

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-038-948-8

seq_documentation_block:

Sequence 8, Application US/08038948
Patent No. 5641487
GENERAL INFORMATION:
APPLICANT: DEAN, JURRIEN
TITLE OF INVENTION: CONTRACEPTIVE VACCINE BASED ON
NUMBER OF SEQUENCES: 14
TITLE OF INVENTION: ALLOIMMUNIZATION WITH ZONA PELLUCIDA POLYPEPTIDES
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
STREET: 1100 New York Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/038,948
FILING DATE: 26-MAR-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/930,462
FILING DATE: 20-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/364,379
FILING DATE: 12-JUN-1989
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, Watson T.
REGISTRATION NUMBER: 26,581
REFERENCE/DOCKET NUMBER: 99152/E-266-88/2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 713 amino acids


```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 713
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: mouse
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL LINE:
; ORGANELLE:
; FEATURE:
; NAME/KEY: ZP2
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: mouse ZP2 protein
;
US-08-453-952-6

alignment_scores:
    Quality: 48.50      Length: 38
    Ratio: 2.553      Gaps: 2
    Percent Similarity: 50.000      Percent Identity: 34.211

alignment_block:
US-09-684-016-48411 x US-08-453-952-6 ..
Align seg 1/1 to: US-08-453-952-6 from: 1 to: 713

      8 TGTAAATGTGAGATGCTTACACACCTGATGACCTAATGGCGCAATGCGA 57
      ||| ||||| ||| :: ||::||| ||||| ::
      353 CysHiscysgluserProvalSerIleaspIleucysalagin...As 368
      58 GGGCTGCACGTGCTGTAAGTTTCACATGTAAGTGATCTACTCTCA 107
      :||| :||::|
      368 pGlyPheMetasphe.....GluV 375

108 TTATTCACACAA 121
::||| |||
375 alTyrSerHisGln 379

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-862-903-6

seq_documentation_block:
; Sequence 6, Application US/08862903
; Patent No. 5916768
; GENERAL INFORMATION:
; APPLICANT: DEAN, JURRIEN
; TITLE OF INVENTION: CONTRACEPTIVE VACCINE
; TITLE OF INVENTION: BASED ON ALLOIMUNIZATION WITH ZONA PELLUCIDA
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/862,903
; FILING DATE: 30-May-1995
; CLASSIFICATION: 424
```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/038,948
; FILING DATE: 26-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,462
; FILING DATE: 20-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/364,379
; FILING DATE: 12-JUN-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AUTH
; REGISTRATION NUMBER: 36,434
; REFERENCE/DOCKET NUMBER: 2026-4032 USA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 713
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: mouse
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE:
; CELL LINE:
; ORGANELLE:
; FEATURE:
; NAME/KEY: ZP2
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: mouse ZP2 protein
;
US-08-862-903-6

alignment_scores:
    Quality: 48.50      Length: 38
    Ratio: 2.553      Gaps: 2
    Percent Similarity: 50.000      Percent Identity: 34.211

alignment_block:
US-09-684-016-48411 x US-08-862-903-6 ..
Align seg 1/1 to: US-08-862-903-6 from: 1 to: 713

      8 TGTAAATGTGAGATGCTTACACACCTGATGACCTAATGGCGCAATGCGA 57
      ||| ||||| ||| :: ||::||| ||||| ::
      353 CysHiscysgluserProvalSerIleaspIleucysalagin...As 368
      58 GGGCTGCACGTGCTGTAAGTTTCACATGTAAGTGATCTACTCTCA 107
      :||| :||::|
      368 pGlyPheMetasphe.....GluV 375

108 TTATTCACACAA 121
::||| |||
375 alTyrSerHisGln 379

seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-549-515-6

seq_documentation_block:
; Sequence 6, Application US/08549515
; Patent No. 6054123
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M
; APPLICANT: Klein, Michel H
; TITLE OF INVENTION: Haemophilus Influenzae
```

```

: TITLE OF INVENTION: Dimethylsulphoxide Reductase Enzyme
:
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Slim & McBurney
: STREET: Suite 701, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/549,515
: FILING DATE: 27-OCT-1995
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Stewart, Michael I
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-522
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 205 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-08-549-515-6

```

alignment_scores:

Quality:	48.00	Length:	27
Ratio:	3.429	Gaps:	1
Percent Similarity:	51.852	Percent Identity:	33.333

alignment_block:

```

US-09-684-016-48411 x US-08-549-515-6  ..
Align seg 1/1 to: US-08-549-515-6 from: 1 to: 205

```

```

2 AGGATGTGTAATGTGAGATGCTTACAACTGATGACCTA..... 43
|||||
103 ArgTYrCYshISmetAlaCYsProTYrAspAlaPrOGlnTYrAspAlaG1 119
44 .....ATGGCGCAATGCGAGGCGCTGC 64
|||||
119 nLYsgLYhISmetThrLYscYsAspGLYcYs 129

```

seq_name: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:US-08-549-515-12

seq_documentation_block:

```

: Sequence 12, Application US/08549515
: Patent No. 6034123
: GENERAL INFORMATION:
: APPLICANT: Loosmore, Sheena M
: APPLICANT: Kleiu, Michel H
: TITLE OF INVENTION: Haemophilus Influenzae
: TITLE OF INVENTION: Dimethylsulphoxide Reductase Enzyme
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Slim & McBurney
: STREET: Suite 701, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible

```

```

: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/549,515
: FILING DATE: 27-OCT-1995
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Stewart, Michael I
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-522
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 205 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
: US-08-549-515-12

```

alignment_scores:

Quality:	48.00	Length:	27
Ratio:	3.429	Gaps:	1
Percent Similarity:	51.852	Percent Identity:	33.333

alignment_block:

```

US-09-684-016-48411 x US-08-549-515-12  ..
Align seg 1/1 to: US-08-549-515-12 from: 1 to: 205

```

```

2 AGGATGTGTAATGTGAGATGCTTACAACTGATGACCTA..... 43
|||||
103 ArgTYrCYshISmetAlaCYsProTYrAspAlaPrOGlnTYrAspAlaG1 119
44 .....ATGGCGCAATGCGAGGCGCTGC 64
|||||
119 nLYsgLYhISmetThrLYscYsAspGLYcYs 129

```

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-249-112-4

seq_documentation_block:

```

: Sequence 4, Application US/08249112
: Patent No. 5527703
: GENERAL INFORMATION:
: APPLICANT: Gully, Joseph F.
: APPLICANT: Liu, Ken K.
: APPLICANT: Arena, Joseph P.
: TITLE OF INVENTION: DNA ENCODING GLUTAMATE GATED CHLORIDE
: TITLE OF INVENTION: CHANNELS
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Wallen, John W.
: STREET: 126 E. Lincoln Ave., P.O. Box 2000
: CITY: Rahway
: STATE: New Jersey
: COUNTRY: U.S.A.
: ZIP: 07065
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/249,112
: FILING DATE: 25-MAY-1994
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Wallen, John W.
: REGISTRATION NUMBER: 35,403
: REFERENCE/DOCKET NUMBER: 19194

```

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (908) 594-3905
 TELEFAX: (908) 594-4720
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 487 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-249-112-4

alignment_scores:
 Quality: 48.00 Length: 28
 Ratio: 2.526 Gaps: 1
 Percent Similarity: 67.857 Percent Identity: 39.286

alignment_block:
 US-09-684-016-48411/rev x US-08-249-112-4 ..

Align seg 1/1 to: US-08-249-112-4 from: 1 to: 487

104 AAGATAGTATCACCCTTACATGTAAGAACTTACAGTCAGTCAGCCCTCG 55
 ||||| :::::::::::|||||||:::||||
 152 LysileuTySerSerArgIleSerleuThrSerSerCysPrometar 168

54 CATTCGC.....CCATTAGTCATCAGGTTGT 27
 |:::| ||||| ::|||::|::|
 168 gLeuGlInleuTyPrProleuAspTyrgInSerCys 179

seq_name: /cgn2_6/plodata/2/1aa/PCTUS_COMB.pep:PCT-US95-06556-4

seq_documentation_block:

Sequence 4, Application PC/TUS9506556
 GENERAL INFORMATION:
 APPLICANT: Cully, Doris F.
 APPLICANT: Arena, Joseph P.
 APPLICANT: Liu, Ken K.
 APPLICANT: Vassiliadis, Demetrios
 TITLE OF INVENTION: DNA ENCODING GLUTAMATE GATED CHLORIDE
 NUMBER OF INVENTION CHANNELS
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Wallen, John W.
 STREET: 126 E. Lincoln Ave., P.O. Box 2000
 CITY: Rahway
 STATE: New Jersey
 COUNTRY: U.S.A.
 ZIP: 07065
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/06556
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/249,112
 FILING DATE: 25-MAY-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Wallen, John W.
 REGISTRATION NUMBER: 35,403
 REFERENCE/DOCKET NUMBER: 19194
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (908) 594-3905
 TELEFAX: (908) 594-4720
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 487 amino acids
 TYPE: amino acid

STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US95-06556-4

alignment_scores:
 Quality: 48.00 Length: 28
 Ratio: 2.526 Gaps: 1
 Percent Similarity: 67.857 Percent Identity: 39.286

alignment_block:

US-09-684-016-48411/rev x PCT-US95-06556-4 ..

Align seg 1/1 to: PCT-US95-06556-4 from: 1 to: 487

104 AAGATAGTATCACCCTTACATGTAAGAACTTACAGTCAGTCAGCCCTCG 55
 ||||| :::::::::::|||||||:::||||
 152 LysileuTySerSerArgIleSerleuThrSerSerCysPrometar 168

54 CATTCGC.....CCATTAGTCATCAGGTTGT 27
 |:::| ||||| ::|||::|::|
 168 gLeuGlInleuTyPrProleuAspTyrgInSerCys 179

seq_name: /cgn2_6/plodata/2/1aa/6B_COMB.pep:US-09-211-704A-7

seq_documentation_block:

Sequence 7, Application US/09211704A
 Patent No. 6271014
 GENERAL INFORMATION:
 APPLICANT: de Saint-Vis, Blandine Marie
 APPLICANT: Fossiez, Francois
 APPLICANT: Caux, Christophe
 APPLICANT: Lebecque, Serge J.F.
 TITLE OF INVENTION: Mammalian Proteinases; Related Reagents
 NUMBER OF INVENTION: and Methods
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: DNAX Research Institute
 STREET: 901 California Avenue
 CITY: Palo Alto
 STATE: California
 COUNTRY: USA
 ZIP: 94304-1104
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/211,704A
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 09/005,263
 FILING DATE: 09-JAN-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Ching, Edwin P.
 REGISTRATION NUMBER: 34,090
 REFERENCE/DOCKET NUMBER: SF0781K
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (650)852-9196
 TELEFAX: (650)496-1200
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 519 amino acids
 TYPE: amino acid
 STRANDEDNESS: not relevant
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-09-211-704A-7

alignment_scores: Quality: 48.00 Length: 36
 Ratio: 2.526 Gaps: 1
Percent Similarity: 52.778 Percent Identity: 33.333

alignment_block:

US-09-684-016-48411/rev x US-09-211-704A-7 ..

Align seg 1/1 to: US-09-211-704A-7 from: 1 to: 519

```
116 TCGAATTAATGAGATAGTATACCTTACATGTGAACTTACAGCTCA 67
||||| : : : : : : : : : : : : : : : : : : : : : : : :
49 TTPASHLYAARGASINLEUSERTPARGVAL.....AT 59
66 GTGACGCCCTGCATGCGCCATTAAGTCATCAAGGTTGTAAGCATCTC 17
: ||||| : : : : : : : : : : : : : : : : : : : : : :
59 gthrphepRoArgAspSerProLeuGlyHisAspThrValArgAlaLeu 76
16 ACATTTC 9
|||
76 eLTYrTYr 78
```

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-115-954-4

seq_documentation_block:

```
; Sequence 4, Application US/09115954B
; Patent No. 6200776
; GENERAL INFORMATION:
; APPLICANT: Boron, Walter F
; APPLICANT: Brill, Antoine M
; APPLICANT: Khandoudi, Nassirah
; APPLICANT: Martin, Xavier
; APPLICANT: Jupe, Steven C
; APPLICANT: Rawlings, Christopher J
; APPLICANT: Doe, Trudy R
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GH-30409
; CURRENT FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: EP97401714.7
; EARLIER FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 714
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-115-954-4
```

alignment_scores: Quality: 48.00 Length: 12
 Ratio: 4.800 Gaps: 0
Percent Similarity: 83.333 Percent Identity: 50.000

alignment_block:

US-09-684-016-48411 x US-09-115-954-4 ..

Align seg 1/1 to: US-09-115-954-4 from: 1 to: 714

```
5 TATTGTAATGTGAGATGCTTACACCGTCATGAC 40
||||| : : : : : : : : : : : : : : : : : : : : : :
305 TYrCYsArgCYsThrLeuProGluAsnProAsnAsn 316
```

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-357-598-12

seq_documentation_block:

```
; Sequence 12, Application US/08357598
; Patent No. 5705625
; GENERAL INFORMATION:
; APPLICANT: Civin, Curt I.
; APPLICANT: Small, Donald
; TITLE OF INVENTION: NOVEL PROTEIN TYROSINE KINASE, JAK3
```

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 4225 Executive Square, Suite 1400

CITY: La Jolla

STATE: CA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/357,598

FILING DATE: 15-DEC-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Haile, Lisa A.

REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: 07265/033001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619/678-5070

TELEFAX: 619/678-5099

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 846 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-357-598-12

alignment_scores:

Quality: 48.00 Length: 36
Ratio: 2.400 Gaps: 2
Percent Similarity: 55.556 Percent Identity: 38.889

alignment_block:

US-09-684-016-48411/rev x US-08-357-598-12 ..

Align seg 1/1 to: US-08-357-598-12 from: 1 to: 846

```
105 GAAGATTAAGTACCTTACATGTGAACTTACAGTCAGCGAGCC.. 58
: : : : : : : : : : : : : : : : : : : : : :
312 AspAspSerPheLeuThrCys.....ValGlnProLe 323
57 .....TCGCATGCCCATTAAGTCATCAGGTTGTAGGCA 21
: : : : : : : : : : : : : : : : : : : : : :
323 uGlyProAspTYrLysGlyCysLeuIleArgProGlyPheLeuValGlyL 340
20 TCTCACAAT 13
: : : : :
340 euserHis 342
```

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-09-003-289-12

seq_documentation_block:

```
; Sequence 12, Application US/09003289
; Patent No. 5916792
; GENERAL INFORMATION:
; APPLICANT: Civin, Curt I.
; APPLICANT: Small, Donald
; TITLE OF INVENTION: NOVEL PROTEIN TYROSINE KINASE, JAK3
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/003,289
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/357,598
FILING DATE: 15-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/033001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 846 amino acids
TYPE: amino acid
STRANDEDNESS: No. 5916792 Relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-003-289-12
```

```
alignment_scores:
Quality: 48.00 Length: 36
Ratio: 2.400 Gaps: 2
Percent Similarity: 55.556 Percent Identity: 38.889
```

alignment_block:

US-09-684-016-48411/rev x US-09-003-289-12 ..

Align seg 1/1 to: US-09-003-289-12 from: 1 to: 846

```
105 GAAGATAAGTATCACCCTTACATGTGAACCTTACAGTCAGTCAGGCC.. 58
:::|||||: 111111111
312 AspAspSerPheLeuThrCys.....ValGlnProLe 323
57 .....TCGCATTGCCCATTTAGGTATCAGGCTTGAAGCA 21
:::|||||: 111111111
323 uGlyProAspTyrIysGlyCysLeuIleArgProGlyPheLeuValGlyL 340
20 TCTCACAT 13
:::|||||:
340 euserHis 342

seq_name: /cgn2.6/ptodata/2/1aa/PCTUS_COMB.pep:PCT-US95-16435-12

seq_documentation_block:
; Sequence 12, Application PC/TUS9516435
; GENERAL INFORMATION:
; APPLICANT: The Johns Hopkins University School of Medicine
; TITLE OF INVENTION: NOVEL PROTEIN TYROSINE KINASE, JAK3
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/16435
```

```
FILING DATE: 15-DEC-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/033001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 846 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-16435-12
```

```
alignment_scores:
Quality: 48.00 Length: 36
Ratio: 2.400 Gaps: 2
Percent Similarity: 55.556 Percent Identity: 38.889
```

alignment_block:

US-09-684-016-48411/rev x PCT-US95-16435-12 ..

Align seg 1/1 to: PCT-US95-16435-12 from: 1 to: 846

```
105 GAAGATAAGTATCACCCTTACATGTGAACCTTACAGTCAGTCAGGCC.. 58
:::|||||: 111111111
312 AspAspSerPheLeuThrCys.....ValGlnProLe 323
57 .....TCGCATTGCCCATTTAGGTATCAGGCTTGAAGCA 21
:::|||||: 111111111
323 uGlyProAspTyrIysGlyCysLeuIleArgProGlyPheLeuValGlyL 340
20 TCTCACAT 13
:::|||||:
340 euserHis 342
```

seq_name: /cgn2.6/ptodata/2/1aa/GB_COMB.pep:US-09-115-954-8

```
seq_documentation_block:
; Sequence 8, Application US/09115954B
; Patent No. 6200776
; GENERAL INFORMATION:
; APPLICANT: Boron, Walter F
; APPLICANT: Brill, Antoine M
; APPLICANT: Khandoudi, Nassirah
; APPLICANT: Martin, Xavier
; APPLICANT: Jupe, Steven C
; APPLICANT: Rawlings, Christopher J
; APPLICANT: Doe, Trudy R
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GH-30409
; CURRENT APPLICATION NUMBER: US/09/115,954B
; CURRENT FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: EP97401714.7
; EARLIER FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 1032
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-115-954-8

alignment_scores:
Quality: 48.00 Length: 12
Ratio: 4.800 Gaps: 0
Percent Similarity: 83.333 Percent Identity: 50.000
```

alignment_block:

US-09-684-016-48411 x US-09-115-954-8 ..

Align seg 1/1 to: US-09-115-954-8 from: 1 to: 1032

```
5 TATGTGAATGTCAGATGCCTTACAAACCTGATGAC 40
|||||:|||||:|||||:|||||:|||||:|||||:
623 TycysArGcysThrLeuProGluAsnProAsnAsn 634
```

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-115-954-2

seq_documentation_block:

```
; Sequence 2, Application US/09115954B
; Patent No. 6200776
; GENERAL INFORMATION:
; APPLICANT: Boron, Walter F
; APPLICANT: Brill, Antoine M
; APPLICANT: Khandoudi, Nassirah
; APPLICANT: Martin, Xavier
; APPLICANT: Jupe, Steven C
; APPLICANT: Rawlings, Christopher J
; APPLICANT: Doe, Trudy R
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GH-30409
; CURRENT APPLICATION NUMBER: US/09/115,954B
; CURRENT FILING DATE: 1998-07-15
; EARLIER APPLICATION NUMBER: EP97401714.7
; EARLIER FILING DATE: 1997-07-16
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1044
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-115-954-2
```

alignment_scores:

Quality:	48.00	Length:	12
Ratio:	4.800	Gaps:	0
Percent Similarity:	83.333	Percent Identity:	50.000

alignment_block:

US-09-684-016-48411 x US-09-115-954-2 ..

Align seg 1/1 to: US-09-115-954-2 from: 1 to: 1044

```
5 TATGTGAATGTCAGATGCCTTACAAACCTGATGAC 40
|||||:|||||:|||||:|||||:|||||:|||||:
635 TycysArGcysThrLeuProGluAsnProAsnAsn 646
```

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-352-990-18

seq_documentation_block:

```
; Sequence 18, Application US/09352990
; Patent No. 6255090
; GENERAL INFORMATION:
; APPLICANT: Farnoczu, Layo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Plant Aminoacyl-trna Synthetase
; FILE REFERENCE: BB-1191
; CURRENT APPLICATION NUMBER: US/09/352,990
; CURRENT FILING DATE: 1999-07-14
; EARLIER APPLICATION NUMBER: 60/092,866
; EARLIER FILING DATE: July 15, 1998
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 18
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Glycine max
```

US-09-352-990-18

alignment_scores:

Quality:	47.00	Length:	20
Ratio:	3.133	Gaps:	0
Percent Similarity:	75.000	Percent Identity:	45.000

alignment_block:

US-09-684-016-48411/rev x US-09-352-990-18 ..

Align seg 1/1 to: US-09-352-990-18 from: 1 to: 400

```
104 AAGATAGATATCACCTTACATGTGAACCTTACCACTGACAGCCCTCG 55
|||||:|||||:|||||:|||||:|||||:|||||:
4 LysMetSerValSerHisPheAlaValLeuSerSerCysCysProAr 20
54 CATTGCCCCA 45
|||||:|||||:
20 glauAlaPro 23
```

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-445-586-9

seq_documentation_block:

```
; Sequence 9, Application US/08445586
; Patent No. 5627050
; GENERAL INFORMATION:
; APPLICANT: Takeshita, Sunao
; APPLICANT: Ito, Toshimi
; APPLICANT: Otawara-Hamamoto, Yokio
; APPLICANT: Amann, Egon
; TITLE OF INVENTION: Bone-Related Sulfatase-Like Protein and
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/445,586
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/111,887
; FILING DATE: 26-AUG-1993
; APPLICATION NUMBER: JP 230030/92
; FILING DATE: 28-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 324034/92
; FILING DATE: 03-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Forman, David S.
; REGISTRATION NUMBER: 33,694
; REFERENCE/DOCKET NUMBER: 02481.1322-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 583 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-445-586-9
```


Ratio: 3.357 Gaps: 0
Percent Similarity: 82.353 Percent Identity: 41.176

alignment_block:

US-09-684-016-48411/rev x US-08-484-494-14 ..

Align seg 1/1 to: US-08-484-494-14 from: 1 to: 583

62 ACCCGTCGCAATGGCCATTAGTCATCAGGCTTGTAAAGCATCTCACAT 13
:::||||::: ||:::|||| ::||| ||:::||||:
470 ThrProAsnPhesapnroValIglySerAsnGlyCysPheAlaThrHisVa 486
12 T 12
486 I 486

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-345-212-14

seq_documentation_block:

Sequence 14, Application US/08345212
Patent No. 5932211
GENERAL INFORMATION:
APPLICANT: Wilson, Peter J
APPLICANT: Morris, Charles P
APPLICANT: Anson, Donald S
APPLICANT: Occhiodoro, Teresa
APPLICANT: Bielicki, Julie
APPLICANT: Clements, Peter R
APPLICANT: Hopwood, John J
TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF
TITLE OF INVENTION: IDURONATE 2-SULFATASE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/345, 212
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 991,973
FILING DATE: 17-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8416Z
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4366
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 583 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-345-212-14

alignment_scores:
Quality: 47.00 Length: 17
Ratio: 3.357 Gaps: 0
Percent Similarity: 82.353 Percent Identity: 41.176

alignment_block:

US-09-684-016-48411/rev x US-08-345-212-14 ..

Align seg 1/1 to: US-08-345-212-14 from: 1 to: 583

62 ACCCGTCGCAATGGCCATTAGTCATCAGGCTTGTAAAGCATCTCACAT 13
:::||||::: ||:::|||| ::||| ||:::||||:
470 ThrProAsnPhesapnroValIglySerAsnGlyCysPheAlaThrHisVa 486
12 T 12
486 I 486

seq_name: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:US-09-249-003-14

seq_documentation_block:

Sequence 14, Application US/09249003
Patent No. 6153188
GENERAL INFORMATION:
APPLICANT: Wilson, Peter J
APPLICANT: Morris, Charles P
APPLICANT: Anson, Donald S
APPLICANT: Occhiodoro, Teresa
APPLICANT: Bielicki, Julie
APPLICANT: Clements, Peter R
APPLICANT: Hopwood, John J
TITLE OF INVENTION: GLYCOSYLATION VARIANTS OF
TITLE OF INVENTION: IDURONATE 2-SULFATASE
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Scully, Scott, Murphy & Presser
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: USA
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/249, 003
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/991,973
FILING DATE: 17-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S
REGISTRATION NUMBER: 31,346
REFERENCE/DOCKET NUMBER: 8416Z
TELECOMMUNICATION INFORMATION:
TELEPHONE: 516-742-4366
TELEFAX: 516-742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 583 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-249-003-14

alignment_scores:
Quality: 47.00 Length: 17
Ratio: 3.357 Gaps: 0
Percent Similarity: 82.353 Percent Identity: 41.176
alignment_block:

372 CysSerGlnProTyrProSerGlyAspSerCysAlaGluProAla 388

...